(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number
WO 02/008260 A3

(51) International Patent Classification[7]: C12N 15/52, 9/10, C12Q 1/68, G01N 33/574, C07K 16/40, A61K 38/43, 39/395, 48/00

(21) International Application Number: PCT/US01/23439

(22) International Filing Date: 26 July 2001 (26.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/220,967	26 July 2000 (26.07.2000)	US
60/251,669	6 December 2000 (06.12.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/220,967 (CIP)
Filed on	26 July 2000 (26.07.2000)
US	60/251,669 (CIP)
Filed on	6 December 2000 (06.12.2000)

(71) Applicants *(for all designated States except US)*: **STANFORD UNIVERSITY** [US/US]; Suite 35, 900 Welch Road, Stanford, CA 94305-1850 (US). **APPLIED GENOMICS, INC.** [—/US]; 325 East Middlefield Road, Mountain View, CA 94043 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BOTSTEIN, David** [US/US]; Stanford University, Suite 35, 900 Welch Road, Stanford, CA 94305-1850 (US). **BROWN, Patrick, O.** [US/US]; 76 Peter Coutts Circle, Stanford, CA 94305 (US). **PEROU, Chuck** [US/US]; The University of North Carolina, At Chapel Hill, Chapel Hill, NC 94305 (US). **ROSS, Douglas** [US/US]; 1437 Cabrillo Avenue, Burlingame, CA 94010 (US). **SEITZ, Rob** [US/US]; 2611 Elderdale Drive, Huntsville, AL 35763 (US).

(74) Agent: **GERBER, Monica, R.**; Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
17 October 2002

(48) Date of publication of this corrected version:
20 March 2003

(15) Information about Correction:
see PCT Gazette No. 12/2003 of 20 March 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/008260 A3

(54) Title: BSTP-ECG1 PROTEIN AND RELATED REAGENTS AND METHODS OF USE THEREOF

(57) Abstract: The invention provides polypeptides (BSTP-ECG1) and polynucleotides which identify and encode BSTP-ECG1 and fragments and variants thereof. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. In addition, the invention provides methods for treating or preventing disorders of cell proliferation, particularly breast cancer, by administering a pharmaceutical composition comprising a polypeptide, polynucleotide, or antibody of the invention. The invention also provides methods of classifying diseases, particularly breast cancer, by detecting expression of BSTP-ECG1 or of a polynucleotide encoding BSTP-ECG1, and of providing diagnostic, prognostic, and/or predictive information for a patient based on the detection and/or measurement of BSTP-ECG1 or of a polynucleotide encoding BSTP-ECG1.

BSTP-ECG1 PROTEIN AND RELATED REAGENTS AND METHODS OF USE THEREOF

GOVERNMENT SUPPORT

The U.S. Government has a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of Grant No. NIH CA 77097 awarded by the National Cancer Institute.

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to provisional applications U.S.S.N 60/220,967, filed July 26, 2000, and U.S.S.N. 60/251,669, filed December 6, 2000, which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

A major challenge of cancer treatment is to target specific therapies to distinct tumor types in order to maximize efficacy and minimize toxicity. A related challenge lies in the attempt to provide accurate diagnostic, prognostic, and predictive information. At present, breast tumors are described with the tumor-node-metastasis (TNM) system. This system, which uses the size of the tumor, the presence or absence of tumor in regional lymph nodes, and the presence or absence of distant metastases, to assign a stage to the tumor is described in the American Joint Committee on Cancer: AJCC Cancer Staging Manual. Philadelphia, Pa: Lippincott-Raven Publishers, 5th ed., 1997, pp 171-180, and further discussion is found in Harris, JR: "Staging of breast carcinoma" in Harris, J.R., Hellman, S., Henderson, I.C., Kinne D.W. (eds.): *Breast Diseases*. Philadelphia, Lippincott, 1991. The assigned stage is used as a basis for selection of appropriate therapy and for prognostic purposes. In addition to the TNM parameters, morphologic appearance is used to further classify tumors and thereby aid in selection of appropriate therapy. However, this approach has serious limitations. Tumors with similar histopathologic appearance can exhibit

significant variability in terms of clinical course and response to therapy. For example, some tumors are rapidly progressive while others are not. Some tumors respond readily to hormonal therapy or chemotherapy while others are resistant.

Assays for cell surface markers, e.g., using immunohistochemistry, have provided means for dividing certain tumor types into subclasses. For example, one factor considered in prognosis and in treatment decisions for breast cancer is the presence or absence of the estrogen receptor (ER) in tumor samples. ER-positive breast cancers typically respond much more readily to hormonal therapies such as tamoxifen, which acts as an anti-estrogen in breast tissue, than ER-negative tumors. Though useful, these analyses are subject to variability and provide only a very crude basis for tumor classification. Therefore, there exists a need for improved methods for classifying tumors.

Mutation or dysregulation of any of a large number of genes contributes to the development and progression of cancer as discussed in Hanahan, D. and Weinberg, R., The Hallmarks of Cancer, *Cell*, 100, 57-70, 2000. Genes that play a role in cancer can be divided into a number of broad classes including oncogenes, tumor suppressor genes, and genes that regulate apoptosis. Oncogenes such as *ras* typically encode proteins whose activities promote cell growth and/or division, a function that is necessary for normal physiological processes such as development, tissue regeneration, and wound healing. However, inappropriate activity or expression of oncogenes can lead to the uncontrolled cell proliferation that is a feature of cancer. Tumor suppressor genes such as *rb* act as negative regulators of cell proliferation. Loss of their activity, e.g., due to mutations or decreased expression at the level of mRNA or protein, can lead to unrestrained cell division. A number of familial cancer syndromes and inherited susceptibility to cancer are believed to be caused by mutations in tumor suppressor genes. Apoptosis, or programmed cell death, plays important roles both in normal development and in surveillance to eliminate cells whose survival may be deleterious to the organism, e.g., cells that have acquired DNA damage. Many chemotherapeutic agents are believed to work by activating the endogenous apoptosis pathway in tumor cells.

Although a substantial number of genes have been implicated as playing important roles in cancer, the factors responsible for the phenotypic diversity of

tumors remain largely unknown. In particular, understanding of the underlying differences in gene expression that may contribute to tumor phenotype is limited. Understanding the differences in gene expression between normal and cancerous tissue and between different tumors of the same tissue type is of significant diagnostic, prognostic, and therapeutic utility. There is therefore a need for the identification of genes exhibiting differential expression in tumors. In particular, there is a need for the identification of additional genes and proteins that can be used to classify tumors, especially genes and proteins that can provide diagnostic, prognostic, and/or predictive information in cancer. There is also a need for antibodies and other reagents for the detection and measurement of such genes and proteins.

Most of the commonly used chemotherapeutic agents act relatively nonselectively. Rather than specifically killing tumor cells, these agents target any dividing cell, resulting in a variety of adverse effects. In addition, current therapeutic strategies are of limited efficacy, and the mortality rate of breast cancer remains high. There is therefore a need for the identification of additional genes and proteins that can be used as targets for the treatment of cancer. There is also a need for antibodies and other reagents that can modulate, regulate, or interact with these genes and proteins to provide new method of treatment for cancer.

SUMMARY OF THE INVENTION

The present invention relates to the identification of genes of particular import in diagnosis, prognostication and/or therapeutic intervention in breast cancer and other tumors based on their expression profile in human breast tumor samples, their expression in other tissue and normal tissue samples, and in cell lines as assessed using cDNA microarrays. In particular, the genes are identified based on their differential expression across tumor samples.

The invention provides a substantially purified polypeptide and fragments thereof that are encoded by an RNA molecule that is differentially expressed in human breast tumor samples and cell lines. The polypeptide is referred to as BSTP-ECG1. Thus in one aspect the invention provides a substantially purified polypeptide whose

amino acid sequence comprises the amino acid sequence set forth in SEQ ID NO:1. The invention also provides polypeptides possessing homology to the polypeptide having the sequence of SEQ ID NO:1 or to fragments of this polypeptide, wherein the polypeptides are significantly similar to the polypeptide of SEQ ID NO:1. In certain embodiments of the invention the definition of "significantly similar" can vary, as described further below. In certain embodiments of the invention a significantly similar polypeptide has one or more amino acid substitutions, deletions, and/or additions with respect to the sequence of SEQ ID NO:1. In certain embodiments of the invention the polypeptides are expression products of human genes. When referring to polypeptides or polynucleotides whose sequence comprises the sequence set forth in a SEQ ID NO:X, the set of polypeptides or polynucleotides includes those polypeptides or polynucleotides having the particular sequence set forth in the SEQ ID NO:X in addition to other polypeptides or polynucleotides including the sequence of SEQ ID NO:X.

In another aspect, the invention provides a substantially isolated and purified polynucleotide encoding the polypeptide of SEQ ID NO:1. In particular, the invention provides a substantially isolated and purified polynucleotide whose sequence comprises the sequence of SEQ ID NO:2. The invention further provides a substantially isolated and purified polynucleotide whose sequence comprises the sequence of SEQ ID NO:3 and also provides substantially isolated and purified polynucleotides whose sequence comprises the sequence of SEQ ID NO:4 or SEQ ID NO:5. The invention also provides a polynucleotide encoding a polypeptide possessing significant similarity to the polypeptide of SEQ ID NO:1, where "significantly similar" is defined below. The invention further provides a polynucleotide having a sequence that is complementary to the sequence of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5. In addition, the invention provides a polynucleotide having a sequence that is complementary to a polypeptide encoding a polypeptide possessing significant similarity to the polypeptide of SEQ ID NO:1.

In another aspect, the invention provides an isolated and purified polynucleotide that hybridizes under stringent conditions to a polynucleotide encoding a polypeptide comprising or having the amino acid sequence set forth in SEQ ID

NO:1. In certain embodiments of the invention the polynucleotide encodes at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 1. In particular, the invention provides an isolated and purified polynucleotide that hybridizes under stringent conditions to a polynucleotide having the sequence set forth in SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5, or fragments of either of these sequences. The invention further provides an isolated an purified polynucleotide that hybridizes under stringent conditions to a polynucleotide encoding a polypeptide having significant similarity to a polypeptide comprising or having the amino acid sequence set forth in SEQ ID NO:1. The invention also provides polynucleotides that hybridize under moderately stringent conditions to the foregoing polynucleotides.

In another aspect, the invention provides a substantially purified oligonucleotide that includes a region of nucleotide sequence that hybridizes to at least 8 consecutive nucleotides of sense or antisense sequence of a nucleotide sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5. The invention also provides a substantially purified oligonucleotide that includes a region of nucleotide sequence that hybridizes to at least 8 consecutive nucleotides of sense or antisense sequence of a nucleotide sequence that encodes the polypeptide of SEQ ID NO:1. In certain embodiments of the invention the oligonucleotide is labeled, e.g., with a fluorescent moiety, enzyme, enzyme substrate, or radioisotope, in order to facilitate detection of the oligonucleotide. The oligonucleotide can be used, e.g., as a probe or a primer, to detect the level of a polynucleotide encoding BSTP-ecg1 in cells and/or tissues, e.g., tissues isolated from a patient. The oligonucleotide can also be used to detect mutations in the gene encoding BSTP-ECG1 and/or to detect amplification or altered expression of the gene encoding BSTP-ECG1, in cells and/or tissues. In certain embodiments of the invention the oligonucleotide is attached to an oligonucleotide microarray. In certain embodiments of the invention the oligonucleotide has a sequence capable of binding specifically with an RNA molecule that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 so as to prevent appropriate processing, transport, or translation of the RNA molecule.

In another aspect the invention provides vectors, e.g. plasmids, containing a polynucleotide encoding a polypeptide comprising the sequence of SEQ ID NO:1. The invention also provides vectors, e.g., plasmids, comprising a polynucleotide encoding a polypeptide possessing significant similarity to the polypeptide of SEQ ID NO:1. In certain embodiments of the invention the polynucleotide comprises the polynucleotide sequence of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5. In certain embodiments of the invention the vectors contain genetic control elements operably linked to the polynucleotide, wherein the genetic control elements direct transcription of the polynucleotide. In certain embodiments of the invention the vectors and genetic control elements are adapted for expression of the polynucleotide in a bacterial cell, a yeast cell, an insect cell, or a mammalian cell. The invention further provides host cells, e.g., bacterial, yeast, insect, and mammalian cells containing an expression vector containing the polynucleotide encoding the polypeptide having the sequence of SEQ ID NO:1. The invention further provides host cells, e.g., bacterial, yeast, insect, and mammalian cells containing an expression vector comprising a polynucleotide encoding a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1.

In another aspect, the invention provides an antibody that specifically bind to a polypeptide whose sequence comprises the amino acid sequence of SEQ ID NO:1. The invention further provides antibodies that specifically bind to a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. The invention further provides methods of detecting a polypeptide whose sequence comprises the amino acid sequence of SEQ ID NO:1. The invention further provides methods of detecting a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. The polypeptides can be detected in a variety of different contexts. For example, the polypeptides can be detected in lysates or extracts derived from cells or tissues, in culture medium, in substantially intact cells or tissue samples (e.g., biopsy specimens), and/or in the blood, urine, serum, ascites, or other body fluids or secretions of a subject.

In another aspect, the invention provides a nonhuman transgenic organism comprising a nonnative DNA molecule encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or an ortholog of such a polypeptide, and

including genetic control elements sufficient to direct transcription of the DNA molecule in at least a subset of the organism's cells. The invention further provides a nonhuman transgenic organism comprising a nonnative DNA molecule encoding a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 and including genetic control elements sufficient to direct transcription of the DNA molecule in at least a subset of the organism's cells. In yet another aspect, the invention provides a nonhuman organism in which a native DNA sequence encoding a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 is deleted, e.g., by homologous recombination.

In another aspect, the invention provides methods for detecting a polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1, in a biological sample. The sample may comprise cells, tissue, body fluid, etc., isolated from a subject. The sample may be processed in a variety of ways prior to application of the methods. For example, the sample may be subjected to purification, reverse transcription, amplification, etc. One such method comprises steps of: (a) hybridizing a nucleic acid complementary to the polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1, to at least one nucleic acid in the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex indicates the presence of a polynucleotide encoding the polypeptide in the biological sample. A second such method comprises steps of: (a) hybridizing a nucleic acid encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1, to at least one nucleic acid complementary to at least one nucleic acid in the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex indicates the presence of a polynucleotide encoding the polypeptide in the biological sample.

In yet another aspect, the invention provides kits for detecting a polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1,

or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. The kits can comprise a polynucleotide that hybridizes specifically to the polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 and, optionally, other materials such as suitable buffers, indicators (e.g., fluorophores, chromophores or enzymes providing same), controls (e.g., an appropriate polynucleotide of this invention), controls, and directions for using the kit.

In another aspect, the invention provides methods for detecting a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. One such method comprises steps of: (a) contacting the biological sample with an antibody that specifically binds to the polypeptide of SEQ ID NO: 1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1; and (b) determining whether the antibody specifically binds to the sample, the binding being an indication that the sample contains the polypeptide. The invention also provides kits for performing these methods. The kits can comprise an antibody (preferably a monoclonal antibody) that binds to the polypeptide and, optionally, other materials such as suitable buffers, indicators (e.g., fluorophores, chromophores or enzymes providing same), controls (e.g., a polypeptide of this invention) and directions for using the kit.

In another aspect, the invention provides methods for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. In one embodiment the method includes the steps of providing a cell that expresses the polypeptide or polypeptide fragment, e.g., a cell containing an expression vector containing a polynucleotide encoding the polypeptide or polypeptide fragment operably linked to genetic control elements that direct transcription of the polynucleotide; maintaining the cell under conditions wherein the polypeptide is produced; harvesting the polypeptide from the cell and/or the culture medium; and, optionally, purifying the polypeptide.

In another aspect, the invention provides methods of predicting whether an individual is at risk for a condition featuring inappropriate cell division, e.g., cancer. One such method comprises the step of determining whether there exists, within a cell and/or tissue of an individual, a mutation in a gene encoding a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1, or whether there exists, within a cell and/or tissue of an individual, a mutation in a regulatory sequence for such a gene. A second such method comprises the step of determining whether an individual expresses a particular allele or variant of a gene comprising a nucleotide sequence set forth in SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5, such allele or variant being present within the general population and inherited from either of the individual's parents rather than constituting a *de novo* mutation within the organism. A third such method comprises the step of determining whether there exists, within a cell and/or tissue of an individual, inappropriate expression of a polynucleotide encoding a polypeptide having an amino acid sequence set forth in SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. A fourth such method comprises the step of determining whether there exists, within a cell and/or tissue of an individual, inappropriate expression of a polypeptide having an amino acid sequence set forth in SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1.

In another aspect, the invention provides methods of classifying a disease, particularly of classifying tumors. One such method includes steps of: (a) obtaining cells or tissue from a site of disease; (b) detecting a polynucleotide encoding a polypeptide having a sequence selected from the group consisting of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1, or a complement of such a polynucleotide; and (c) assigning the disease to one of a set of predetermined categories based on detection of the polynucleotide. The method can further comprise the step of providing diagnostic, prognostic, or predictive information based on the category assigned in the assigning step.

In certain embodiments of the invention the method is used to classify breast tumors. In certain embodiments of the invention the method is based on detecting

expression of a single gene, e.g., a gene encoding the polypeptide of SEQ ID NO:1. Detecting expression of such a gene may comprise detecting the polynucleotide sequence set forth in SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5. Detecting expression may comprise measuring either a relative or absolute level of expression. In certain embodiments of the invention the method is based on an assessment of the expression of multiple polynucleotides as described further in copending U.S. Patent Application "Reagents and Methods for Use in Managing Breast Cancer, filed July 26, 2001 and in U.S. Provisional Patent Application Ser. No. 60/220,967, filed July 26, 2000. These applications are referred to herein as "the gene subset application". The multiple polynucleotides can include all of the polynucleotides disclosed therein or any subset thereof.

In another aspect, the invention provides a method of classifying a tumor by detection of the polypeptide of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 in cells and/or tissue samples obtained from the tumor or from elsewhere in the body (e.g., in blood, urine, ascites, other body fluids or secretions or excretions). In a preferred embodiment the method is used to classify breast tumors. In certain embodiments the method is based on the detection of a single polypeptide, e.g., the polypeptide of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. However, in other embodiments the method is based on an assessment of the expression of multiple polypeptides. The multiple polypeptides can include all of the polypeptides disclosed in the gene subset application mentioned above or any subset thereof. In certain embodiments the level of expression (either relative or absolute) of the polypeptide is measured. In other embodiments the pattern of expression of the polypeptide within cells or within a tissue sample is assessed. Regardless of the method by which a tumor is assigned to a predetermined category, the assignment may be used as a basis to provide diagnostic, prognostic, and/or predictive information to the patient having the tumor.

In another aspect, the invention provides a microarray for use in classifying tumors, comprising a polynucleotide whose sequence comprises or is complementary to that set forth in SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO:5, or whose sequence is of sufficient length to specifically bind to such a sequence under

the microarray hybridization conditions employed. Such conditions may be those described in the Examples, or any conditions appropriate for the particular microarray and detection technology employed. The invention further provides a microarray for use in classifying tumors comprising a polynucleotide, e.g., a cDNA or an oligonucleotide, capable of binding specifically to a polynucleotide encoding the polypeptide having the amino acid sequence set forth in SEQ ID NO: 1, or capable of binding specifically to a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1, or complementary to such polynucleotides.

In another aspect, the invention provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence set forth in SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. In certain embodiments of the invention the pharmaceutical composition further preferably comprises a pharmaceutically acceptable carrier.

In yet another aspect, the invention provides a pharmaceutical composition comprising a substantially purified antibody that binds to a polypeptide having an amino acid sequence set forth in SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. In certain embodiments of the invention the pharmaceutical composition further preferably comprises a pharmaceutically acceptable carrier. In certain embodiments the antibody is modified, e.g., by attaching a toxic molecule thereto.

The invention provides methods for identifying modulators of the expression or activity of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. The invention further provides agonists and antagonists capable of modulating the expression or activity of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. The invention provides pharmaceutical compositions including such modulators and methods of use thereof for the treatment or prevention of cancer, particularly breast cancer.

In another aspect, the invention provides a method for the treatment or prevention of cancer comprising the step of administering to an individual in need thereof, a pharmaceutical composition comprising a polypeptide having an amino acid

sequence comprising the sequence of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. In another aspect, the invention provides a method for the treatment or prevention of cancer comprising the step of administering to an individual in need thereof, a pharmaceutical composition comprising an antibody or a modified antibody that binds to a polypeptide having an amino acid sequence set forth in SEQ ID NO:1, or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1.

In another aspect, the invention provides a method of treating or preventing a tumor comprising steps of: (i) providing an individual in need of treatment or prevention of a tumor, (ii) administering a compound that enhances the level or activity of a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1. In certain embodiments of the invention the compound is provided as a component of a pharmaceutical composition. The invention also includes such a pharmaceutical composition.

In another aspect, the invention provides methods of inhibiting growth of a cell comprising enhancing the level or activity of a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 in the cell. The cell can be a normal cell or a tumor cell, e.g., a breast tumor cell. According to certain of the methods the level of the polypeptide is enhanced by overexpressing the polypeptide in the cell, e.g., by introducing an expression vector or other nucleotide sequence (e.g., DNA, RNA, or modified nucleotides, etc.) that encodes the polypeptide into the cell.

In another aspect, the invention provides a method of increasing cell growth comprising:

decreasing the level or activity of a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 in the cell. Various methods for decreasing the level of a polypeptide in a cell are known in the art. Such methods include, for example, the introduction or expression of antisense nucleic acids or double-stranded RNA (RNA-mediated interference) into the cell. Another such method is introduction or expression of dominant negative polypeptides into the cell.

According to one aspect, the invention provides a method of classifying a tumor comprising the steps of (i) providing a tumor sample, (ii) detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the sample; and (iii) classifying the tumor as belonging to a tumor subclass based on the results of the detecting step. The detecting step may comprise detecting the polypeptide. A variety of detection techniques may be employed including, but not limited to, immunohistochemical analysis, ELISA assay, antibody arrays, or detecting modification of a substrate by the polypeptide.

In certain embodiments of the methods the tumor is a breast tumor and the tumor subclass is a luminal tumor subclass. The methods may further comprise providing diagnostic, prognostic, or predictive information based on the classifying step. Classifying may include stratifying the tumor (and thus stratifying a subject having the tumor), e.g., for a clinical trial. The methods may further comprise selecting a treatment based on the classifying step. In clinical research, stratification is the process or result of describing or separating a patient population into more homogeneous subpopulations according to specified criteria. Stratifying patients initially rather than after the trial is frequently preferred, e.g., by regulatory agencies such as the U.S. Food and Drug Administration that may be involved in the approval process for a medication. In some cases stratification may be required by the study design. Various stratification criteria may be employed in conjunction with detection of expression of one or more basal marker genes. Commonly used criteria include age, family history, lymph node status, tumor size, tumor grade, etc. Other criteria including, but not limited to, tumor aggressiveness, prior therapy received by the patient, ER and/or PR positivity, Her2neu status, p53 status, various other biomarkers, etc., may also be used. Stratification is frequently useful in performing statistical analysis of the results of a trial.

In another aspect, the invention provides a method of testing a subject comprising the steps of (i) providing a sample isolated from a subject, (ii) detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the sample, and (iii) providing diagnostic, prognostic, or predictive information based on the detecting step. The detecting step may comprise detecting the polypeptide. Detection may be performed using any appropriate technique including, but not

limited to, immunohistochemistry, ELISA assay, protein array, or detecting modification of a substrate by the polypeptide.

The sample may comprise mRNA, in which case the detecting step may comprise hybridizing the mRNA or cDNA or RNA synthesized from the mRNA to a microarray or detecting mRNA transcribed from the gene or detecting cDNA or RNA synthesized from mRNA transcribed from the gene. In any of the above methods, the sample may be a blood sample, a urine sample, a serum sample, an ascites sample, a saliva sample, a cell, and a portion of tissue.

According to another aspect, the invention provides a method of testing a compound or a combination of compounds for activity against tumors comprising steps of (i) obtaining or providing tumor samples taken from subjects who have been treated with the compound or combination of compounds, wherein the tumors fall within a tumor subclass, (ii) comparing the response rate of tumors that fall within the tumor subclass and have been treated with the compound with the overall response rate of tumors that have been treated with the compound or combination of compounds or with the response rate of tumors that do not fall within the subclass and have been treated with the compound or combination of compounds and (iii) identifying the compound or combination of compounds as having selective activity against tumors in the tumor subclass if the response rate of tumors in the subclass is greater than the overall response rate or the response rate of tumors that do not fall within the subclass. In certain embodiments of the invention the tumors are breast tumors. In certain embodiments of the invention the tumor subclass is a luminal tumor subclass. The tumors may be classified according to any of the inventive classification methods described above. In certain embodiments of the invention the classification is based on expression of the polypeptide of SEQ ID NO:1.

The invention further provides a method of testing a compound or a combination of compounds for activity against tumors comprising steps of (i) treating subjects in need of treatment for tumors with the compound or combination of compounds, (ii) comparing the response rate of tumors that fall within a tumor subclass with the overall response rate of tumors or with the response rate of tumors that do not fall within the subclass, and (iii) identifying the compound or combination of compounds as having selective activity against tumors in the tumor subclass if the

response rate of tumors in the subclass is greater than the overall response rate or the response rate of tumors that do not fall within the subclass. The method may further comprise various additional steps. For example, the method may comprise steps of (i) providing tumor samples from subjects in need of treatment for tumors, (ii) determining whether the tumors fall within a tumor subclass, and (iii) stratifying the subjects based on the results of the determining step prior to performing the treating step. The method may further comprise the steps of (i) providing tumor samples from subjects in need of treatment for tumors, (ii) detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the samples, and (iii) stratifying the subjects based on the results of the detecting step prior to performing the treating step.

In addition, the invention includes a method of testing a compound or a combination of compounds for activity against tumors comprising steps of (i) treating subjects in need of treatment for tumors with the compound or combination of compounds or with an alternate compound, wherein the tumors fall within a tumor subclass, (ii) comparing the response rate of tumors treated with the compound or combination of compounds with the response rate of tumors treated with the alternate compound; and (iii) identifying the compound or combination of compounds as having superior activity against tumors in the tumor subclass, as compared with the alternate compound, if the response rate of tumors treated with the compound or combination of compounds is greater than the response rate of tumors treated with the alternate compound. The method may further comprise various additional steps. For example, the method may comprise steps of (i) providing tumor samples from subjects in need of treatment for tumors, (ii) determining whether the tumors fall within a tumor subclass, and (iii) stratifying the subjects based on the results of the determining step prior to performing the treating step. The method may further comprise the steps of (i) providing tumor samples from subjects in need of treatment for tumors, (ii) detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the samples, and (iii) stratifying the subjects based on the results of the detecting step prior to performing the treating step.

In certain embodiments of the invention the alternate compound is a compound approved by the U.S. Food and Drug administration for treatment of

tumors. The invention also provides a method of treating a subject comprising steps of (i) identifying a subject as having a tumor in a luminal tumor subclass, and (ii) administering to the subject a compound identified according to any of the inventive methods for identifying a subject.

BRIEF DESCRIPTION OF THE DRAWING

Figure 1A presents the sequence of BSTP-ECG1 (SEQ ID NO:1).

Figure 1B presents the polynucleotide sequence of an open reading frame that encodes BSTP-ECG1 (SEQ ID NO: 2)

Figures 1C and 1D present the sequences of two polynucleotides that encode BSTP-ECG1 (SEQ ID NO:3 and SEQ ID NO: 4). These polynucleotides are cDNAs that represent multiple mRNA isoforms arising due to alternate 3' polyadenylation sites.

Figure 2 presents the consensus sequence derived from I.M.A.G.E. clones 161484, 48805, 1276329, 1343900, and 1560906 (SEQ ID NO: 5)

Figure 3 presents an alignment of BSTP-ECG1 with a number of related proteins identified in GenBank.

Figure 4 presents a sequence map in which the predicted transmembrane domain of BSTP-ECG1 (amino acids 66 - 115) is highlighted in gray.

Figure 5A presents a Kyte-Doolittle hydrophobicity plot for BSTP-ECG1.

Figure 5B presents a prediction of transmembrane regions and orientation for BSTP-ECG1 obtained using the program TMpred.

Figure 5C presents a prediction of transmembrane helices for BSTP-ECG1 produced using a Hidden Markov Model.

Figure 6A presents a Northern blot showing expression of *BST-ECG1* in various cell lines.

Figure 6B presents a longer exposure of the image in Figure 6A.

BRIEF DESCRIPTION OF THE TABLES

The tables contain the numerical data corresponding to microarray images. Other tables provide additional information or list the individual genes in the various gene subsets.

Table 1 is a master data table for the 65 microarray experiments performed on individual tumor samples, in which rows represent I.M.A.G.E. clones that identify approximately 1753 genes whose expression varied by at least a factor of 4 and columns represent individual microarray experiments. The first 50 pages of the table consist of a reference list in which a descriptive name for each clone (where such a name exists) appears in the column entitled Name, followed by the Genbank accession number for the clone. Each row in the reference list contains a number in the first column that numerically identifies the column. In the subsequent data portion of the table (pages 1 – 392), each row is similarly identified by a number in the first column so that the name and Genbank accession number for the clone for which data appears in that row may be determined by consulting the reference list. In the data portion of the table, the column headings in the first row identify the tumor samples. Each data cell in the table represents the measured Cy5/Cy3 fluorescence ratio at the corresponding target element on the appropriate array. Empty cells indicate insufficient or missing data. All ratio values are log transformed (base 2) to treat inductions or repressions of identical magnitude as numerically equal but with opposite sign.

Table 2 is a master data table for the 19 microarray experiments performed on cell line samples, in which rows represent I.M.A.G.E. clones that identify approximately 1753 genes whose expression varied by at least a factor of 4 and columns represent individual microarray experiments. This table contains only a data portion, in which the column headings in the first row identify the cell lines. Each row in the table is

identified by a number which appears in the first column. The same reference list that forms part of Table 1 may be consulted to determine the name and Genbank accession number for the clone for which data appears in that row. Each data cell in the table represents the measured Cy5/Cy3 fluorescence ratio at the corresponding target element on the appropriate array. Empty cells indicate insufficient or missing data. All ratio values are log transformed (base 2) to treat inductions or repressions of identical magnitude as numerically equal but with opposite sign.

Table 3 presents a listing and description of the 11 cell lines used to create the common reference sample.

Table 4 presents a complete listing of the 84 experimental samples that were assayed versus the common reference sample. The table includes a list of alternate names (in the column entitled Sample ID/old name) for the same tumors. The alternate names are used to identify the tumor samples in certain contexts, and the table allows conversion between the two sets of names.

Table 5 lists the tumors used in the experiments described herein, along with clinical and pathological information about each tumor/patient.

Table 6 is a master data table for the 84 microarray experiments performed on individual tumor, tissue, and cell line samples, in which rows represent I.M.A.G.E. clones that identify the 496 genes in the intrinsic gene set, and columns represent individual microarray experiments. The first 15 pages of the table consist of a reference list in which a descriptive name for each clone (where such a name exists) appears in the column entitled Name, followed by the Genbank accession number for the clone. Each row in the reference list contains a number in the first column that numerically identifies the column. In the subsequent data portion of the table (pages 1 – 91), each row is similarly identified by a number in the first column so that the name and Genbank accession number for the clone for which data appears in that row may be determined by consulting the reference list. In the data portion of the table, the column headings in the first row identify the tumor samples. Each data cell in the

table represents the measured Cy5/Cy3 fluorescence ratio at the corresponding target element on the appropriate array. Empty cells indicate insufficient or missing data. All ratio values are log transformed (base 2) to treat inductions or repressions of identical magnitude as numerically equal but with opposite sign.

Table 7 is a listing of the 374 clones that identify genes selected for the epithelial enriched gene set including Genbank accession numbers.

Table 8 is a listing of the clones that identify genes that comprise the luminal subset including Genbank accession numbers.

Tables 9-1 and 9-2 are listings of the two groups of clones that identify genes that comprise the basal subset including Genbank accession numbers.

Table 10 is a listing of the clones that identify genes that comprise the *ErbB2* subset including Genbank accession numbers.

Table 11 is a listing of the clones that identify genes that comprise the endothelial gene subset including Genbank accession numbers.

Table 12 is a listing of the clones that identify genes that comprise the stromal/fibroblast gene subset including Genbank accession numbers.

Table 13 is a listing of the clones that identify genes that comprise the B-cell gene subset including Genbank accession numbers.

Table 14 is a listing of the clones that identify genes that comprise the adipose-enriched/normal breast gene subset including Genbank accession numbers.

Table 15 is a listing of the clones that identify genes that comprise the macrophage gene subset including Genbank accession numbers.

Table 16 is a listing of the clones that identify genes that comprise the T-cell gene subset including Genbank accession numbers.

In Table 1, the Genbank accession number for each clone appears in the column entitled "Name", following a brief descriptive name for the gene identified by the clone, where available. In some cases the descriptive name is a number corresponding to an I.M.A.G.E. clone ID number. As is well known and accepted in the art, the Genbank accession number represents a means of definitively identifying a particular clone, since Genbank accession numbers will be maintained permanently or, if changed, the change will be accomplished in such a manner as to allow unambiguous correlation between any new numbering system and the numbering system currently in use.

Note that Tables 1, 2, and 6 are provided for purposes of presenting the clone identifications and the data that was used to perform hierarchical clustering analysis, and that the format of the tables may not correspond exactly with the format required by software developed for the analysis of the data. Appropriate format will, in general, depend upon the particular computer program. See, for example, the Web site http://genome-www.stanford.edu/~sherlock/tutorial.html for discussion of the appropriate format for one particular analysis program.

In Tables 7 – 16, each entry identifies a clone. The first portion of each entry is a brief descriptive name for the gene identified by the clone. The Genbank accession number for the clone appears on the last line of the entry for that clone.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

DEFINITIONS

To facilitate understanding of the description of the invention, the following definitions are provided. It is to be understood that, in general, terms not otherwise defined are to be given their meaning or meanings as generally accepted in the art.

Agonist: As used herein, the term "agonist" refers to a molecule that increases or prolongs the duration of the effect of a polypeptide or a nucleic acid. Agonists may

include proteins, nucleic acids, carbohydrates, lipids, small molecules, ions, or any other molecules that modulate the effect of the polypeptide or nucleic acid. An agonist may be a direct agonist, in which case it is a molecule that exerts its effect by binding to the polypeptide or nucleic acid, or an indirect agonist, in which case it exerts its effect via a mechanism other than binding to the polypeptide or nucleic acid (e.g., by altering expression or stability of the polypeptide or nucleic acid, by altering the expression or activity of a target of the polypeptide or nucleic acid, by interacting with an intermediate in a pathway involving the polypeptide or nucleic acid, etc.)

Antagonist: As used herein, the term "antagonist" refers to a molecule that decreases or reduces the duration of the effect of a polypeptide or a nucleic acid. Antagonists may include proteins, nucleic acids, carbohydrates, or any other molecules that modulate the effect of the polypeptide or nucleic acid. An antagonist may be a direct antagonist, in which case it is a molecule that exerts its effect by binding to the polypeptide or nucleic acid, or an indirect antagonist, in which case it exerts its effect via a mechanism other than binding to the polypeptide or nucleic acid (e.g., by altering expression or stability of the polypeptide or nucleic acid, by altering the expression or activity of a target of the polypeptide or nucleic acid, by interacting with an intermediate in a pathway involving the polypeptide or nucleic acid, etc.)

Corresponding to: In general, the phrase "corresponding to" has its commonly accepted meaning indicating, typically, a relationship between two entities, etc. For example, a mRNA corresponds to a gene if the mRNA is transcribed from the gene. A protein corresponds to a gene if the protein is translated from an mRNA transcribed from the gene. A cDNA corresponds to an mRNA if the cDNA is synthesized by reverse transcription of the mRNA. In addition, and without limitation, as used herein, an mRNA corresponds to a clone on a microarray when the mRNA (or cDNA derived therefrom) hybridizes specifically (under the experimental conditions described) to the clone or to its complement, e.g., when the sequence of the mRNA (or cDNA derived therefrom) and the sequence of the clone are sufficiently complementary to one another for specific hybridization to occur. Similarly, a gene corresponds to a clone on a microarray when mRNA transcribed from the gene

corresponds to the clone. Note that it is not necessary that the entire mRNA, cDNA, etc. hybridize with the clone or vice versa. For example, the mRNA or cDNA may be longer or shorter than the clone. The clone may be longer or shorter than the mRNA or cDNA. Either or both of the mRNA/cDNA and the clone may contain one or more stretches of sequence that is/are not contained within the corresponding nucleic acid.

Diagnostic information: As used herein, diagnostic information or information for use in diagnosis is any information that is useful in determining whether a patient has a disease or condition and/or in classifying the disease or condition into a phenotypic category or any category having significance with regards to the prognosis of or likely response to treatment (either treatment in general or any particular treatment) of the disease or condition. Similarly, diagnosis refers to providing any type of diagnostic information, including, but not limited to, whether a subject is likely to have a condition (such as a tumor), information related to the nature or classification of a tumor, information related to prognosis and/or information useful in selecting an appropriate treatment. Selection of treatment may include the choice of a particular chemotherapeutic agent or other treatment modality such as surgery, radiation, etc., a choice about whether to withhold or deliver therapy, etc.

Differential expression: A gene exhibits differential expression at the RNA level if its RNA transcript varies in abundance between different samples in a sample set. A gene exhibits differential expression at the protein level, if a polypeptide encoded by the gene varies in abundance between different samples in a sample set. In the context of a microarray experiment, differential expression generally refers to differential expression at the RNA level.

Gene: For the purposes of the present invention, the term "gene" has its meaning as understood in the art. However, it will be appreciated by those of ordinary skill in the art that the term "gene" has a variety of meanings in the art, some of which include gene regulatory sequences (e.g., promoters, enhancers, etc.) and/or intron sequences, 3' untranslated regions, etc., and others of which are limited to coding sequences. It will further be appreciated that definitions of "gene" include references to nucleic

acids that do not encode proteins but rather encode functional RNA molecules such as tRNAs. For the purpose of clarity we note that, as used in the present application, the term "gene" generally refers to a portion of a nucleic acid that encodes a protein; the term may optionally encompass regulatory sequences. This definition is not intended to exclude application of the term "gene" to non-protein coding expression units but rather to clarify that, in most cases, the term as used in this document refers to a protein coding nucleic acid.

Gene product or expression product: A gene product or expression product is, in general, an RNA transcribed from the gene or a polypeptide encoded by an RNA transcribed from the gene.

Homology: The term "homology" refers to a degree of similarity between two or more nucleic acid sequences or between two or more amino acid sequences. As is well known in the art, given any nucleotide or amino acid sequence, homologous sequences may be identified by searching databases (e.g., GenBank, EST [expressed sequence tag]databases, GST [gene sequence tag] databases, GSS [genome survey sequence] databases, organism sequencing project databases) using computer programs such as BLASTN for nucleotide sequences and BLASTP, gapped BLAST, and PSI-BLAST for amino acid sequences. These programs are described in Altschul, SF, et al., Basic local alignment search tool, *J. Mol. Biol.*, 215(3): 403-410, 1990, Altchul, SF and Gish, W, *Methods in Enzymology*, and Altschul, SF, et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402, 1997. In addition to identifying homologous sequences, the programs mentioned above typically provide an indication of the degree of homology. Determining the degree of identity or homology that exists between two or more amino acid sequences or between two or more nucleotide sequences can also be conveniently performed using any of a variety of other algorithms and computer programs known in the art. Discussion and sources of appropriate programs may be found, for example, in Baxevanis, A., and Ouellette, B.F.F., *Bioinformatics : A Practical Guide to the Analysis of Genes and Proteins*, Wiley, 1998; and Misener, S. and Krawetz, S. (eds.), *Bioinformatics Methods and*

Protocols (Methods in Molecular Biology, Vol. 132), Humana Press, 1999.

Operably linked: The term "operably linked" refers to a relationship between two nucleic acid sequences wherein the expression of one of the nucleic acid sequences is controlled by, regulated by, modulated by, etc. the other nucleic acid sequence. For example, a promoter is operably linked with a coding sequence if the promoter controls transcription of the coding sequence. Preferably a nucleic acid sequence that is operably linked to a second nucleic acid sequence is covalently linked, either directly or indirectly, to such a sequence, although any effective three-dimensional association is acceptable.

Prognostic information and predictive information: As used herein the terms prognostic information and predictive information are used interchangeably and somewhat informally to refer to any information that may be used to foretell any aspect of the course of a disease or condition either in the absence or presence of treatment. Such information may include, but is not limited to, the average life expectancy of a patient, the likelihood that a patient will survive for a given amount of time (e.g., 6 months, 1 year, 5 years, etc.), the likelihood that a patient will be cured of a disease, the likelihood that a patient's disease will respond to a particular therapy (wherein response may be defined in any of a variety of ways). Prognostic and predictive information are included within the broad category of diagnostic information.

Sample: As used herein, a sample obtained from a subject may include, but is not limited to, any or all of the following: a cell or cells, a portion of tissue, blood, serum, ascites, urine, saliva, and other body fluids, secretions, or excretions. The term "sample" also includes any material derived by processing such a sample. Derived samples may include nucleic acids or proteins extracted from the sample or obtained by subjecting the sample to techniques such as amplification or reverse transcription of mRNA, etc.

Specific binding: As used herein, the term refers to an interaction between a target polypeptide (or, more generally, a target molecule) and a binding molecule such as an antibody, agonist, or antagonist. The interaction is typically dependent upon the presence of a particular structural feature of the target polypeptide such as an antigenic determinant or epitope recognized by the binding molecule. For example, if an antibody is specific for epitope A, the presence of a polypeptide containing epitope A or the presence of free unlabeled A in a reaction containing both free labeled A and the antibody thereto, will reduce the amount of labeled A that binds to the antibody. It is to be understood that specificity need not be absolute. For example, it is well known in the art that numerous antibodies cross-react with other epitopes in addition to those present in the target molecule. Such cross-reactivity may be acceptable depending upon the application for which the antibody is to be used. Thus the degree of specificity of an antibody will depend on the context in which it is being used. In general, an antibody exhibits specificity for a particular partner if it favors binding of that partner above binding of other potential partners. One of ordinary skill in the art will be able to select antibodies having a sufficient degree of specificity to perform appropriately in any given application (e.g., for detection of a target molecule, for therapeutic purposes, etc). It is also to be understood that specificity may be evaluated in the context of additional factors such as the affinity of the binding molecule for the target polypeptide versus the affinity of the binding molecule for other targets, e.g., competitors. If a binding molecule exhibits a high affinity for a target molecule that it is desired to detect and low affinity for nontarget molecules, the antibody will likely be an acceptable reagent for immunodiagnostic purposes. Once the specificity of a binding molecule is established in one or more contexts, it may be employed in other, preferably similar, contexts without necessarily re-evaluating its specificity.

Treating a tumor: As used herein, treating a tumor is taken to mean treating a subject who has the tumor.

Tumor sample: The term "tumor sample" as used herein is taken broadly to include cell or tissue samples removed from a tumor, cells (or their progeny) derived from a

tumor that may be located elsewhere in the body (e.g., cells in the bloodstream or at a site of metastasis), or any material derived by processing such a sample. Derived tumor samples may include nucleic acids or proteins extracted from the sample or obtained by subjecting the sample to techniques such as amplification or reverse transcription of mRNA, etc.

Tumor subclass: A tumor subclass, also referred to herein as a tumor subset or tumor class, is the group of tumors that display one or more phenotypic or genotypic characteristics that distinguish members of the group from other tumors.

Vector: A vector, as used herein, is a nucleic acid molecule that includes sequences sufficient to direct *in vivo* or *in vitro* replication of the molecule. These may either be self-replication sequences or sequences sufficient to direct integration of the vector into another nucleic acid present in a cell (either an endogenous nucleic acid or one introduced into the cell by experimental manipulation), so that the vector sequences are replicated during replication of this nucleic acid. Preferred vectors include a cloning site, at which foreign nucleic acid molecules may be introduced. Vectors may include control sequences that have the ability to direct *in vivo* or *in vitro* expression of nucleic acid sequences introduced into the vector. Such control sequences may include, for example, transcriptional control sequences (e.g., promoters, enhancers, terminators, etc.), splicing control sequences, translational control sequences, etc. Vectors may also include a coding sequence, so that transcription and translation of sequences introduced into the vector results in production of a fusion protein.

I. Overview

The invention is based on the identification of polynucleotides (cDNAs) corresponding to human genes that are differentially expressed in human breast tumor samples, the polypeptides encoded by these polynucleotides, and antibodies raised against these polypeptides. The invention encompasses the use of these polynucleotides, polypeptides, and antibodies as well as compositions containing them, either singly or in combination, in the prediction, diagnosis, treatment, or

prevention of cancer and in the provision of prognostic and predictive information related to cancer.

Nucleic acids encoding BSTP-ECG1 were identified based on expression profiles gathered in a series of cDNA microarray experiments. As described in more detail in Examples 1, 2, and 4, cDNA microarrays each representing the same set of approximately 8100 different human genes were produced. The human cDNA clones used to produce the microarrays contained approximately 4000 named genes, 2000 genes with homology to named genes in other species, and approximately 2000 ESTs of unknown function. An mRNA sample was obtained from each of a set of 84 tissue samples or cell lines. The expression levels of the approximately 8100 genes were measured in each mRNA sample by hybridization to an individual microarray, yielding an expression profile for each gene across the experimental samples. These expression profiles were studied and compared and were used to identify nucleic acids encoding BSTP-ECG1. Although more details will be found in the Examples, a description of cDNA microarray technology and a description of the experimental approach employed in the identification of the polynucleotides that encode BSTP-ECG1 is presented here so that the invention may be better understood. Certain aspects of the invention are then described in detail.

II. cDNA Microarray Technology

cDNA microarrays consist of multiple (usually thousands) of different cDNAs spotted (usually using a robotic spotting device) onto known locations on a solid support, such as a glass microscope slide. After spotting, the cDNAs are usually cross-linked to the support, e.g., by UV irradiation. The cDNAs are typically obtained by PCR amplification of plasmid library inserts using primers complementary to the vector backbone portion of the plasmid or to the gene itself for genes where sequence is known. PCR products suitable for production of microarrays are typically between 0.5 and 2.5 kB in length. Full length cDNAs, expressed sequence tags (ESTs), or randomly chosen cDNAs from any library of interest can be chosen. ESTs are partially sequenced cDNAs as described, for example, in L. Hillier, et al., Generation and analysis of

280,000 human expressed sequence tags, *Genome Research*, 6, 807-828, 1996. The afore-mentioned article is herein incorporated by reference, as are the entire teachings of all other patents and journal articles mentioned herein, for all purposes and not just those related to the particular context in which they are mentioned. The present application also incorporates by reference six U.S. patent applications filed by inventors on July 26, 2001. These applications are entitled "REAGENTS AND METHODS FOR USE IN MANAGING BREAST CANCER", "BSTP-RAS/RERG PROTEIN AND RELATED REAGENTS AND METHODS OF USE THEREOF", "BSTP-CAD PROTEIN AND RELATED REAGENTS AND METHODS OF USE THEREOF", "BASAL CELL MARKERS IN BREAST CANCER AND USES THEREOF", "BSTP-TRANS PROTEIN AND RELATED REAGENTS AND METHODS OF USE THEREOF", "BSTP-5 PROTEINS AND RELATED REAGENTS AND METHODS OF USE THEREOF".

Although some ESTs correspond to known genes, frequently very little or no information regarding any particular EST is available except for a small amount of 3' and/or 5' sequence and, possibly, the tissue of origin of the mRNA from which the EST was derived. As will be appreciated by one of ordinary skill in the art, in general the cDNAs contain sufficient sequence information to uniquely identify a gene within the human genome. Furthermore, in general the cDNAs are of sufficient length to hybridize specifically to cDNA obtained from mRNA derived from a single gene under the hybridization conditions of the experiment.

In a typical microarray experiment, a microarray is hybridized with differentially labeled RNA or DNA populations derived from two different samples. Most commonly RNA (either total RNA or poly A^+ RNA is isolated from cells or tissues of interest and is reverse transcribed to yield cDNA. Labeling is usually performed during reverse transcription by incorporating a labeled nucleotide in the reaction mixture. Although various labels can be used, most commonly the nucleotide is conjugated with the fluorescent dyes Cy3 or Cy5. For example, Cy5-dUTP and Cy3-dUTP can be used. cDNA derived from one sample (representing, for example, a particular cell type, tissue type or growth condition) is labeled with one fluor while cDNA derived from a second sample (representing, for example, a different cell type, tissue type, or growth condition) is labeled with the second fluor.

Similar amounts of labeled material from the two samples are cohybridized to the microarray. In the case of a microarray experiment in which the samples are labeled with Cy5 (which fluoresces red) and Cy3 (which fluoresces green), the primary data (obtained by scanning the microarray using a detector capable of quantitatively detecting fluorescence intensity) are ratios of fluorescence intensity (red/green, R/G). These ratios represent the relative concentrations of cDNA molecules that hybridized to the cDNAs
represented on the microarray and thus reflect the relative expression levels of the mRNA corresponding to each cDNA/gene represented on the microarray.

Each microarray experiment can provide tens of thousands of data points, each representing the relative expression of a particular gene in the two samples. Appropriate organization and analysis of the data is of key importance. Various computer programs have been developed to facilitate data analysis. One basis for organizing gene expression data is to group genes with similar expression patterns together into clusters. A method for performing hierarchical cluster analysis and display of data derived from microarray experiments is described in Eisen, M., Spellman, P., Brown, P., and Botstein, D., Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. USA*, 95: 14863-14868, 1998. As described therein, clustering can be combined with a graphical representation of the primary data in which each data point is represented with a color that quantitatively and qualitatively represents that data point. By converting the data from a large table of numbers into a visual format, this process facilitates an intuitive analysis of the data. Additional information and details regarding the mathematical tools and/or the clustering approach itself may be found, for example, in Sokal, R.R. & Sneath, P.H.A. Principles of numerical taxonomy, xvi, 359, W. H. Freeman, San Francisco,1963; Hartigan, J.A. Clustering algorithms, xiii, 351, Wiley, New York, 1975; Paull, K.D. et al. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J Natl Cancer Inst* 81, 1088-92,1989; Weinstein, J.N. et al. Neural computing in cancer drug development:
predicting mechanism of action. *Science* 258, 447-51, 1992; van Osdol, W.W., Myers, T.G., Paull, K.D., Kohn, K.W. & Weinstein, J.N. Use of the Kohonen self-

organizing map to study the mechanisms of action of chemotherapeutic agents. *J Natl Cancer Inst* 86, 1853-9, 1994; and Weinstein, J.N. *et al.* An information-intensive approach to the molecular pharmacology of cancer. *Science*, 275, 343-9, 1997.

Further details of the experimental methods used in the present invention are found in the Examples. Additional information describing methods for fabricating and using microarrays is found in U.S. Patent No. 5,807,522, which is herein incorporated by reference. Instructions for constructing microarray hardware (e.g., arrayers and scanners) using commercially available parts can be found at http://cmgm.stanford.edu/pbrown/ and in Cheung, V., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., and Childs, G., Making and reading microarrays, *Nature Genetics Supplement*, 21:15-19, 1999, which are herein incorporated by reference. Additional discussions of microarray technology and protocols for preparing samples and performing microrarray experiments are found in, for example, DNA arrays for analysis of gene expression, *Methods Enzymol*, 303:179-205, 1999; Fluorescence-based expression monitoring using microarrays, *Methods Enzymol*, 306: 3-18, 1999; and M. Schena (ed.), DNA Microarrays: A Practical Approach, Oxford University Press, Oxford, UK, 1999. Descriptions of how to use an arrayer and the associated software are found at http://cmgm.stanford.edu/pbrown/mguide/arrayerHTML/ArrayerDocs.html, which is herein incorporated by reference.

III. Experimental Approach of the Invention

The present invention encompasses the realization that genes that are differentially expressed in tumors are of use in tumor classification and are targets for the development of diagnostic and therapeutic agents. Differentially expressed genes are likely to be responsible for the different phenotypic characteristics of tumors. The present invention identifies one such gene. In general, a differentially expressed gene is a gene whose transcript abundance varies between different tumor samples. For example, and without intending to be limiting, the transcript level of a differentially expressed gene may vary by at least fourfold from its average abundance in a given sample set in at least 1 sample, at least two samples, at least three samples. etc. Of course other criteria for differential expression may be employed.

While analysis of multiple genes is of use in developing a robust classification of tumors, each of the differentially expressed genes and their encoded proteins is a target for the development of diagnostic and therapeutic agents. Investigation of variation in individual genes in breast tumors reveals that molecular variation can be related to important features of clinical variation. For example, expression of the estrogen receptor alpha gene (*ESR1*), the *Erb-B2/HER2/neu* oncogene (the target for the monoclonal antibody Herceptin® (Trastuzumab), a recently approved treatment for certain patients with metastatic breast cancer), and the mutational status at the *TP53*, *BRCA1* and *BRCA2* loci have shown that molecular variation can be related to important features of clinical variation. (Discussed, for example, in Osborne, C.K., *et al.*, The value of estrogen and progesterone receptors in the treatment of breast cancer, *Cancer* 46, 2884-2888, 1980; Ingvarsson, S., Molecular genetics of breast cancer progression, *Seminars in Cancer Biology*, 9, 277-288, 1999; Breast Cancer Linkage Consortium, Pathology of familial breast cancer: differences between breast cancers in carriers of *BRCA1* and *BRCA2* mutations and sporadic cases, *Lancet*, 349, 1505-1510, 1997; Anderson, T. I., et al., Prognostic significance of *TP53* alterations in breast carcinoma. *Br J Cancer*, 68, 540-548, 1993 and references cited in these articles). In particular, approximately 60% to 70% of breast tumors express the estrogen receptor, and this expression has been shown to be a favorable prognostic factor (reviewed in Allred, D.C., et al. Prognostic and Predictive Factors in Breast Cancer by Immunohistochemical Analysis, *Modern Pathology*, 11(2), 155-168, 1998). As these examples demonstrate, the study of genetic variation in breast cancer has the potential to contribute to improved classification, diagnosis, and therapy for patients.

As described in more detail in Examples 1, 2, and 4, in order to identify genes that are differentially expressed in breast tumors, cDNA microarrays each representing the same set of approximately 8100 different human genes were produced. The human cDNA clones used to produce the microarrays contained approximately 4000 named genes, 2000 genes with homology to named genes in other species, and approximately 2000 ESTs of unknown function. An mRNA sample was obtained from each of a set of 84 tissue samples or cell lines. The expression levels of the approximately 8100 genes were measured in each mRNA sample by hybridization to an individual microarray, yielding an expression profile for each gene across the

experimental samples. Although more details will be found in the Examples, an overview of the experimental procedure is presented here so that the invention may be better understood.

Variation in patterns of gene expression were characterized in 62 breast tumor samples from 40 different patients, 3 normal breast tissue samples, and 19 samples from 17 cultured human cell lines (one of which was sampled 3 times under different conditions). Twenty of the tumors had been sampled twice, before and after a 16 week course of doxorubicin chemotherapy, and two tumors were paired with a lymph node metastasis from the same patient. The other 18 tumor samples were single samples from individual tumors. A detailed listing of the tumor samples and various characteristics including clinical estrogen receptor and Erb-B2 status as assessed using antibody staining, estrogen receptor and Erb-B2 status as assessed by microarray result, tumor grade, differentiation, survival status and time, age at diagnosis, doxorubicin response, and p53 status is presented in Table 5 of the gene subset application. A listing of the cell lines including description and ATCC (American Tissue Culture Collection) number or reference is presented in Table 3 of the gene subset application. The cell lines provided a framework for interpreting the variation in gene expression patterns seen in the tumor samples and included gene expression models for many of the cell types encountered in tumors.

As described in more detail in Example 2, mRNA was isolated from each sample. cDNA labeled with the fluorescent dye Cy5 was prepared from each experimental sample separately. Fluorescently labeled cDNA, labeled using a second distinguishable dye (Cy3), was prepared from a pool of mRNAs isolated from 11 different cultured cell lines. The pooled mRNA sample served as a reference to provide a common internal standard against which each gene's expression in each experimental sample was measured.

Comparative expression measurements were made by separately mixing Cy5-labeled experimental cDNA derived from each of the 84 samples with a portion of the Cy3-labeled reference cDNA, and hybridizing each mixture to an individual cDNA microarray. The ratio of Cy5 fluorescence to Cy3 fluorescence measured at each cDNA element on the microarray was then quantitatively measured. The use of a common reference standard in each hybridization allowed the fluorescence ratios to be

treated as comparative measurements of the expression level of each gene across all the experimental samples.

A hierarchical clustering method (Eisen, *et al.*, 1998) was used to group genes based on similarity in the pattern with which their expression varied over all experimental samples. The same clustering method was used to group the experimental samples (tissue and cell lines separately) based on the similarity in their patterns of expression. Interpretation of the data obtained from the clustering algorithm was facilitated by displaying the data in the form of tumor and gene dendrograms. In the tumor dendrograms, the pattern and length of the branches reflects the relatedness of the tumor samples with respect to their expression of genes represented on the microarray. Microarray images, and tumor and gene dendrograms are available in Perou, et. al, *Nature*, 2000, and at inventors' Web site (http://genome-www.stanford.edu/molecularportraits/). The similarity of the gene expression profiles of individual tumor samples or groups of tumor samples to one another is inversely related to the length of the branches that connect them. Thus, for example, adjacent tumor samples connected to one another by short vertical branches descending from a common horizontal branch (e.g., tumor samples Norway 48-BE and Norway 48-AF close to the right of the tumor dendrogram) are more closely related to one another in terms of their gene expression profiles than adjacent tumor samples connected to one another by longer vertical branches descending from a common horizontal branch (e.g., tumor samples Norway 100-BE and Norway 100-AF at the left side of the tumor dendrogram). To the extent that the gene expression programs dictate the biological properties and behavior of the tumors and reflect their physiological state and environment, it is expected that the clustering of the tumors reflects phenotypic relationships among them, e.g., tumor samples connected by short horizontal branches (i.e., located in close proximity to one another) are expected to exhibit similar phenotypic features. In the gene dendrograms, the pattern and length of the branches reflects the relatedness of the genes with respect to their expression profiles across the tumor samples. Similarly to the tumor samples, genes connected by short vertical branches are more similar to one another in terms of expression profile than genes connected by longer vertical branches.

The expression patterns of the genes were also displayed using a matrix format, with each row representing all of the hybridization results for a single cDNA element on the array and each column representing the measured expression levels for all genes in a single sample. In this format, tumor samples with similar patterns of expression across the gene set are close to each other along the horizontal dimension. Similarly, genes with similar expression patterns across the set of samples are close to each other along the vertical dimension. To allow the patterns of expression to be visualized, the normalized expression value of each gene was represented by a colored box, using red to represent expression levels greater than the median and green to represent expression levels less than the median. In all array images the brightest red color represents transcript levels at least 16-fold greater than the median, and the brightest green color represents transcript levels at least 16-fold below the median. This display format facilitates comparisons between genes and the recognition of significant patterns. Certain gene subsets of particular interest are indicated by colored bars along the right side of the matrices. These subsets are discussed further in the gene set application.

IV. Identification and characterization of sequences encoding BSTP-ECG1

I.M.A.G.E. clone 161484 was identified based on the expression pattern of its corresponding mRNA among the 84 samples analyzed by microarray hybridization. In particular, transcripts corresponding to clone 161484 varied in abundance by at least 4-fold from their median abundance in the sample set, among at least 3 of the 84 samples. Thus the polynucleotide corresponding to clone 161484 was differentially expressed among the tumor samples, indicating its potential utility for classifying tumors. Numerical data indicating the measured expression of mRNA corresponding to clone 161484 (i.e., mRNA hybridizing to clone 161484) appear in Table 1. In Table 1, information pertaining to clone 161484 is entered under the GenBank accession number of the clone, i.e., H25606. The complete entry on the referral page of Table 1 for clone 161484 is: ESTS, WEAKLY SIMILAR TO W01A11.2 GENE PRODUCT [C.ELEGANS] H25606. In the color matrix representations the expression profile for clone 161484 is identified with the label ESTS, WEAKLY SIMILAR TO W01A11.2 GENE PRODUCT [C.ELEGANS]. The expression profile

and additional data related to clone 161484 are available at Applicants' Web site (http://genome-www.stanford.edu/molecularportraits/) in Supplementary Figure 4. As shown therein, the expression level of mRNA corresponding to clone 161484 varied significantly among tumor samples. For example, mRNA corresponding to clone 161484 was particularly highly expressed in tumor samples Norway 18-BE, Norway 104-BE, Norway 112-BE, Norway 112-AF, and Stanford 14. The relative expression level was also high in tumors Norway 18-AF, Norway 12-AF, and Norway 26-BE, among others. The relative expression level of mRNA corresponding to clone 161484 was particularly low in tumor samples Norway 27-BE and Norway 7-AF and was also low in tumor samples Norway 15-BE, Stanford 38-P, Stanford 38-LN, Norway 56, Norway 16, Stanford 24, Norway 27-AF, New York 1, Norway 39-AF, and Norway 102-BE, among others. As is evident from the discussion above, the presence of mRNA corresponding to clone 161484 reflects the expression of the gene from which the mRNA is transcribed.

A search of GenBank revealed that only a small portion at the 3' end and a small portion at the 5' end of clone 161484 had been sequenced. To confirm the identity of the clone actually used on the array, sequencing runs from the 3' and 5' ends of the clone were performed. As expected, the sequences obtained corresponded to the 3' and 5' sequences in GenBank. Overlapping clones (I.M.A.G.E. clones 48805, 1276329, 1343900, and 1560906) were identified by first searching GenBank for clones homologous to clone 161484 and then searching for additional clones homologous to the clones identified as homologous to clone 161484. A consensus nucleotide sequence (SEQ ID NO: 5) was derived based on sequencing and analysis of overlapping I.M.A.G.E. clones 161484, 48805, 1276329, 1343900, and 1560906. In SEQ ID NO: 5, the abbreviation N stands for any nucleotide. The consensus sequence (SEQ ID NO: 5) was used as input for a search of GenBank with the BLASTX program (which translates a nucleotide sequence in each of the possible six reading frames and then searches for homologous amino acid sequences). The search indicated that a portion of the translated amino acid sequence in one reading frame had homologs in a large number of eukaryotic species including *C. elegans*. An open reading frame (SEQ ID NO: 2) encoding this portion was identified within the

consensus sequence. The predicted amino acid sequence for the polypeptide encoded by this open reading frame is presented as SEQ ID NO: 1.

Thus in one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO: 1. The polypeptide is 388 amino acids in length. A search of the GenBank database using the BLASTP computer program (Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402) performed with this sequence indicated that this polypeptide has at least one homolog in a large number of eukaryotic species, consistent with the information obtained from the BLASTX search described above. Due to the fact that the polypeptide comprising the amino acid sequence presented in SEQ ID NO:1 is highly conserved across eukaryotes, the polypeptide will be referred to herein as BSTP-ECG1 (for Breast Protein - Eukaryotic Conserved Gene 1, and the gene encoding BSTP-ECG1 will be referred to as *BST-ECG1*. While not wishing to be bound by any theory, the fact that ECG1 is so highly conserved across eukaryotes may indicate that it performs an essential cellular function. Figure 3 presents an alignment of BSTP-ECG1 with a number of related proteins identified in GenBank, in which identical amino acids are shaded. GenBank accession numbers are listed before the name of each protein. BSTP-ECG1 is 40-37% identical to the other proteins in this alignment.

Analysis of the BSTP-ECG1 coding sequence using three different techniques indicated the presence of a putative transmembrane domain between amino acids 66 and 115. Figure 4 presents a sequence map of BSTP-ECG1 in which the predicted transmembrane domain is highlighted in gray. Figure 5A presents a Kyte-Doolittle hydrophobicity plot for BSTP-ECG1 (Kyte J and Doolittle RF A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology* 157(6): 105-142, 1982). Figure 5B presents a prediction of transmembrane regions and orientation for BSTP-ECG1 obtained using the program TMpred (K. Hofmann & W. Stoffel (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* 347,166). Figure 5C presents a prediction of transmembrane helices for BSTP-ECG1 produced using a hidden Markov model (Erik

L.L. Sonnhammer, Gunnar von Heijne, and Anders Krogh: A hidden Markov model for predicting transmembrane helices in protein sequences. In *Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology*, p 175-182 Ed J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen. Menlo Park, CA: AAAI Press, 1998.

Northern analysis (see Figures 6A and 6B and Example 10) revealed that *BST-ECG1* mRNA can be detected in a variety of tumor-derived cell lines, including a breast adenocarcinoma cell line (MCF-7). While not wishing to be bound by any theory, this data suggests that in addition to being useful for the diagnosis and/or therapy of breast cancer, the methods and reagents described herein may be useful in the diagnosis and/or therapy of additional tumor types. Northern analysis confirmed the existence of multiple mRNA isoforms encoding BSTP-ECG1, which arise due to alternate 3' polyadenylation sites. The Northern blot showed 2 bands of approximately 1.5 and 2.2 kB. The sequence of a cDNA corresponding to the ~2.2 kB band is presented as SEQ ID NO:3 (2445 nucleotides). The sequence of a cDNA corresponding to the ~1.5 kB band is presented as SEQ ID NO: 4 (1543 nucleotides). In both SEQ ID NO: 3 and SEQ ID NO: 4 the initiation codon begins at nucleotide 228, and the stop codon begins at nucleotide 1392.

The fact that *BST-ECG1* is differentially expressed among breast tumors indicates that the expression level of *BST-ECG1* and/or of its encoded polypeptide, BSTP-ECG1, can be used to distinguish between different subsets of breast tumors. For example, while not wishing to be bound by any theory, the expression level of *BST-ECG1* and/or BSTP-ECG1 may be used, either alone or in combination with other data, to distinguish between tumors falling into phenotypic categories such as a good prognosis category, a poor prognosis category, a nonresponder category (where a different nonresponder category may be defined with respect to each particular therapy), etc. The various categories may be defined in any of a variety of ways and need not be absolute. For example, a good prognosis category may be defined as a category including tumors for which the average survival of patients having tumors in the category is greater than 10 years. As another example, a nonresponder category may be defined as a category including tumors for which the average response rate to a particular therapy is less than 5%, where a "response" can also be defined according

to criteria typically employed in studies such as clinical trials. The expression of *BST-ECG1* may be used to provide prognostic information for breast cancer patients and/or in the selection of appropriate therapy.

Determining the expression of *BST-ECG1* may include measuring mRNA transcribed from *BST-ECG1*, e.g., mRNA comprising a nucleotide sequence set forth in SEQ ID NO:2 or 3 or variants thereof. Determining the expression of *BST-ECG1* may include detecting (qualitatively or quantitatively) a translation product of *BST-ECG1*, e.g., a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1 or a variant thereof.

The discovery of BSTP-ECG1 and the discovery of the differential expression of the gene encoding BSTP-ECG1 satisfy a need in the art by providing compositions useful in the diagnosis, treatment, and prevention of cancer, particularly breast cancer, and by providing methods useful in the classification of cancer and the provision of prognostic information to patients with cancer. Furthermore, BSTP-ECG1 is likely to be a transmembrane protein, indicating that it will likely be accessible to therapeutic agents such as antibodies and/or small molecules. These results suggest that *BST-ECG1* may be a useful gene to target for therapeutic intervention in subsets of breast cancer and a useful gene to distinguish between different subsets of breast cancer.

V. Further Aspects of the Invention

A. Polynucleotides, polypeptides, antibodies, vectors, and host cells

The invention encompasses a polypeptide whose amino acid sequence has or comprises the sequence set forth in SEQ ID NO:1. The invention also encompasses polypeptides possessing significant similarity to BSTP-ecg1, i.e., polypeptides whose sequence possesses signficant similarity to the sequence of SEQ ID NO:1. In certain embodiments of the invention a significantly similar polypeptide has one or more amino acid substitutions, deletions, and/or additions with respect to the sequence of SEQ ID NO:1. In certain embodiments of the invention a significantly similar polypeptide is encoded by a human gene. Definitions of "significantly similar" make reference to the BLAST algorithm and BLOSUM substitution matrix, which are described in Altschul, SF, et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402, 1997 and

Henikoff, S. and Henikoff, J., "Amino acid substitution matrices from protein blocks", *Proc. Natl. Acad. Sci.* 89, 10915-10919, 1992.

In certain embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 25% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 50% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 75% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 90% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 95% of the length of SEQ ID NO:1, or both.

In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared

with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 70 or a % positive greater than 80 encompassing at least 25% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 70 or a % positive greater than 80 encompassing at least 75% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 70 or a % positive greater than 80 encompassing at least 90% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 70 or a % positive greater than 80 encompassing at least 95% of the length of SEQ ID NO:1, or both.

In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 80 or a % positive greater than 90 encompassing at least 25% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 80 or a % positive greater than 90 encompassing at least 75% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if,

when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 80 or a % positive greater than 90 encompassing at least 90% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 80 or a % positive greater than 90 encompassing at least 95% of the length of SEQ ID NO:1, or both.

In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 90 or a % positive greater than 95 encompassing at least 25% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 90 or a % positive greater than 95 encompassing at least 75% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 90 or a % positive greater than 95 encompassing at least 90% of the length of SEQ ID NO:1, or both. In other embodiments of the invention a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 90 or a % positive greater than 95 encompassing at least 95% of the length of SEQ ID NO:1, or both. By "encompassing at least X%

of the length of SEQ ID NO:1" (or, in general, SEQ ID NO:Y) is meant that the length of the portion of SEQ ID NO:1 that is being compared with a potentially similar protein is at least X% of the length of SEQ ID NO:1 (or, in general, SEQ ID NO:Y).

In certain embodiments of the invention a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 includes one or more conservative amino acid substitutions. Examples of conservative substitutions are well known in the art. See, for example, *Biochemistry*, 4th Ed., Stryer, L., *et al.*, W. Freeman and Co., 1995 and U.S. Patent No. 6,015,692. The invention also encompasses variants of the polypeptide of SEQ ID NO:1 and variants of significantly similar polypeptide, wherein the variants have one or more altered or modified amino acids. Alterations and modifications may include the replacement of an L- amino acid with a D-amino acid, or various modifications including, but not limited to, phosphorylation, carboxylation, alkylation, etc.

Certain polypeptides having significant similarity to the polypeptide of SEQ ID NO:1 contain at least one functional or structural characteristic of BSTP-ecg1. For example, certain of the polypeptides contain an epitope that binds an antibody that binds to BSTP-ecg1. Certain of the polypeptides have amino acid sequences that differ by less than 20, less than 10, or less than 5 amino acids from the amino acid sequence of SEQ ID NO: 1. Certain of the polypeptides retain at least one biological activity, structural feature, or immunological activity of BSTP-ECG1.

The invention also encompasses BSTP-ECG1 variants. Certain BSTP-ECG1 variants are at least about 80%, more preferably at least about 90%, and most preferably at least about 95% identical in amino acid sequence to a BSTP-ECG1 amino acid sequence, e.g., the amino acid sequence of SEQ ID NO: 1. Certain variant amino acid sequences differ by less than 20, less than 10, or less than 5 amino acids from the amino acid sequence of SEQ ID NO: 1.

The invention also encompasses fragments of BSTP-ECG1. Preferred BSTP-ECG1 fragments retain at least one biological activity, structural feature, or immunological activity of BSTP-ECG1. In certain embodiments of the invention the fragments are between 5 and 15 amino acids in length. Such fragments are useful, for example, as antigens for the generation of antibodies. In certain embodiments of the

invention the length of the fragments is at least 50%, at least 75%, at least 90%, at least 95%, or at least 99% of the full length of BSTP-ECG1.

The invention also includes polynucleotides that encode BSTP-ECG1. In a particularly preferred embodiment the invention encompasses a polynucleotide comprising the polynucleotide sequence of SEQ ID NO: 2. In another preferred embodiment the invention encompasses a polynucleotide comprising the polynucleotide sequence of SEQ ID NO: 3. In another preferred embodiment the invention encompasses a polynucleotide comprising the polynucleotide sequence of SEQ ID NO: 4. The invention further includes polynucleotides that encode the inventive polypeptide variants described above.

The invention also encompasses a variant of a polynucleotide sequence encoding BSTP-ECG1. Certain variants have at least about 80%, more preferably at least about 90%, and most preferably at least about 95% sequence identity to a polynucleotide sequence encoding BSTP-ECG1. Certain embodiments of the invention include a variant of the polynucleotide sequence of SEQ ID NO: 2 which has at least about 80%, at least about 90%, or at least about 95% sequence identity to the polynucleotide sequence of SEQ ID NO: 2. In another embodiment the invention includes a variant of the polynucleotide sequence of SEQ ID NO: 3 which has at least about 80%, at least about 90%, or at least about 95% sequence identity to the polynucleotide sequence of SEQ ID NO: 3. In another embodiment the invention includes a variant of the polynucleotide sequence of SEQ ID NO:4 which has at least about 80%, at least about 90%, or at least about 95% sequence identity to the polynucleotide sequence of SEQ ID NO:4. Certain variant polynucleotide sequences differ by less than 20, less than 10, or less than 5 nucleotides from the original sequence. The invention further includes polynucleotides that encode the inventive polypeptide variants and fragments described above. Certain polynucleotide variants and fragments encode an amino acid sequence that contains at least one functional or structural characteristic of BSTP-ECG1. Certain polynucleotide fragments comprise at least about 50%, at least about 75%, at least about 80%,at least about 90%, or at least about 95% of the polynucleotide sequence of SEQ ID NO: 2, the polynucleotide sequence of SEQ ID NO: 3, or the polynucleotide sequence of SEQ ID NO: 4.

As is well known in the art, due to the degeneracy of the genetic code (i.e., the fact that in many cases multiple different codons can code for the same amino acid), multiple different polynucleotide sequences encode BSTP-ECG1 of the present invention. The invention encompasses all of the sequences that can be made by substituting alternative codons in accordance with the genetic code. It is noted that such substitution must be made appropriately in view of the reading frame of the polynucleotide. It is further noted that many of these polynucleotide sequences will display little or no homology with the sequences in SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4. In certain embodiments of the invention it may be preferred to employ polynucleotides encoding BSTP-ECG1 having a significantly different codon usage to that found in naturally occurring BSTP-ECG1. For example, if the inventive polynucleotides are to be used to express BSTP-ECG1 in a heterologous system such as a bacterial or yeast expression system, it may be desirable to employ polynucleotides having a codon usage preferred for optimal expression in the heterologous system. Such codon usage preferences are well known in the art. Altering the nucleotide sequence encoding BSTP-ECG1 may have additional uses such as maximizing RNA stability, as it is well known in the art that RNA stability can be affected by the sequence of the RNA.

The invention also includes polynucleotides having a complementary nucleotide sequence to any of the inventive polynucleotides described above. Such complementary polynucleotides are useful as probes, e.g., to detect expression of the inventive polynucleotides at the RNA level. Such complementary polynucleotides are also useful as antisense reagents, to inhibit the expression of the corresponding genes at the protein level, e.g., by interfering with mRNA translation. Inhibiting gene expression has a variety of applications, e.g., it may be used to gain information about the function of the encoded protein. In addition, antisense inhibition of gene expression may be used therapeutically.

The invention encompasses polynucleotides that are able to hybridize to any of the inventive polynucleotide sequences discussed above under various conditions of stringency. In general, a hybridizing polynucleotide will have a sequence either partially or fully complementary with the polynucleotide to which it hybridizes. Hybridization conditions of various stringency are well known in the art and are, in

general, governed by the concentration of reagents such as salts and formamide in the hybridization buffer as well as by the temperature at which hybridization is performed. For example, a stringent hybridization can be performed by use of a hybridization buffer comprising 30% formamide in 0.9M saline/0.09M sodium citrate (SSC) buffer at a temperature of 45° C followed by washing twice with that SSC buffer at 45° C. A moderately stringent hybridization condition could include use of a hybridization buffer comprising 20% formamide in 0.8M saline/0.08M SSC buffer at a temperature of 37° C. followed by washing once with that SSC buffer at 37° C. Further examples of stringent conditions are found in U.S. Patent No. 6,008,337; in Maniatis, T., Sambrook, J. and Fritsch, E., *Molecular Cloning: A Laboratory Manual* (3 Volume Set); and in numerous other sources known to one of ordinary skill in the art. Appropriate stringency conditions to achieve particular degrees of hybridization specificity when using cDNA or oligonucleotide arrays are also well known. The selection of appropriate hybridization conditions will typically be determined by the purpose for which hybridization is to be carried out and is a matter of choice for the practitioner.

The invention further includes oligonucleotides comprising a fragment of a polynucleotide encoding BSTP-ECG1 or comprising a fragment of a polynucleotide complementary to such a polynucleotide. Preferred oligonucleotides are between 6 nucleotides and 60 nucleotides in length, preferably approximately 15 to 30 nucleotides in length, and more preferably between about 20 and 25 nucleotides in length. Such oligonucleotides are useful, for example, as primers in PCR amplification or in hybridization assays including microarray assays.

The invention contemplates the production of any of the polynucleotides or fragments thereof described herein by chemical synthesis, by PCR, or by use of expression vectors including the polynucleotides. Such expression vectors and methods of their use are well known in the art. The inventive polynucleotides and fragments thereof can be produced using an *in vitro* transcription system or within a host cell containing a vector comprising the inventive polynucleotide sequence and appropriate genetic control elements (e.g., enhancers, promoters, terminators) operatively linked to the polynucleotide sequence so as to direct transcription therefrom. *In vitro* transcription systems are well known in the art as are vectors

containing appropriate genetic control elements for directing transcription of an inserted polynucleotide sequence and host cells in which such vectors are maintained. The invention encompasses the production of either DNA or RNA having the sequence of an inventive polynucleotide. One of ordinary skill in the art will be able to select appropriate vectors and synthesis conditions depending upon whether it is desired to produce DNA or RNA. It is noted that the inventive polynucleotides and fragments thereof can also be synthesized entirely through chemical means. Techniques and machines for chemical synthesis of polynucleotides are well known in the art. The polynucleotides can be labeled or conjugated with detectable moieties including radionuclides, enzymes, chromogenic substrates, fluorescent substances, etc., using any of a variety of techniques.

Polynucleotides encoding BSTP-ECG1 can be extended, e.g., to identify upstream elements such as promoters or other regulatory elements, using techniques that are well known in the art. Such techniques are described, for example, in U.S. Patent No. 6,008, 337, which is herein incorporated by reference, and include a variety of PCR-based methods, screening of cDNA libraries, primer extension, etc. Genomic sequence such as introns can also be obtained. Discovery of additional sequence may also be performed using computer-based searches of sequenced human DNA. As is well known, large portions of the human genome have been sequenced, but relatively little information exists as to the structure and organization of much of the sequence. Thus extension of the inventive polynucleotide sequences may be performed by careful examination of previously sequenced genomic DNA. Preferably any predictions based on examination of genomic sequence are verified experimentally, since it is well known in the art that a significant number of errors exist in the genomic sequence and in the predictions (e.g., predictions of genes and open reading frames) based thereon.

It is well known in the art that different cell types, cells at different stages of differentiation, and/or cells within organisms at different developmental stages may express variants of the same gene, e.g., variants derived by alternative splicing. Therefore, multiple different mRNAs corresponding to the same gene may exist. Such mRNAs are transcribed from the same region of genomic DNA but may vary in sequence, usually lacking or containing domains corresponding to introns or regions

at the 5' and/or 3' end of the message. The present invention encompasses such variant polynucleotides and their encoded polypeptides. Variant polynucleotides can be identified and cloned by screening cDNA libraries produced using mRNA from cells of various different types, differentiation states, or from tissues at different developmental stages.

The invention contemplates production of any of the BSTP-ECG1 polypeptides or fragments thereof using any of a variety of techniques including both *in vivo* or *in vitro* synthesis. For example, polynucleotides encoding an inventive polypeptide can be inserted into an expression vector, which can then be introduced into an appropriate host cell (e.g., a bacterial, yeast, insect, or mammalian cell). Thus the invention includes an expression vector comprising a polynucleotide comprising a nucleotide sequence set forth in SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4 or comprising a variant of a nucleotide sequence set forth in SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4. The invention further includes a host cell comprising any of the afore-mentioned vectors. A wide variety of vector/host expression systems are known in the art. In general, vectors contain necessary control and regulatory sequences (e.g., enhancers, promotors, polyadenylation sequence, etc.) operatively linked to an inserted polynucleotide so as to direct expression of the polypeptide in the appropriate host cell. Depending upon the host cell to be employed, appropriate vectors may include phages, viruses, or plasmids. The invention encompasses any available vector/host expression system and specifically includes vectors that direct expression of an inventive polynucleotide, vectors that direct expression of an inventive polypeptide (e.g., expression vectors), in addition to cells and cell lines transformed with such vectors. In certain embodiments of the invention the inventive polypeptide is secreted from a cell transformed with an expression vector, thereby allowing purification from the medium rather than by harvesting the cells. The invention also encompasses the production of inventive polypeptides in cells that have been engineered to express such polypeptides according to the methods described in U.S. Patent No. 6,063,630, which discloses methods of "turning on" an endogenous gene in cells that normally express the gene at low or undetectable levels. Methods for harvesting, isolating, purifying, etc., polypeptides from cells expressing such polypeptides are well known in the art.

BSTP-ECG1 and BSTP-ECG1 variants and fragments thereof can also be produced in animals or plants that are transgenic for a polynucleotide sequence encoding the polypeptide. The invention includes such animals and plants. In addition to their potential use as sources for the inventive polypeptides, transgenic animals may be used to study the function of the inventive polypeptides. Methods for the production of transgenic animals and plants, as well as methods for purifying and harvesting inventive polypeptides from such animals and plants are well known in the art and are within the scope of the invention. The invention also encompasses cells and transgenic animals that have been engineered to lack expression of the inventive polynucleotides. Methods for "knocking out" a gene using the technique of homologous recombination and methods of creating cells and organisms lacking expression of the knocked out gene are well known in the art and are described, for example, in U.S. Patent No. 5,464,764, U.S. Patent No. 5,487,992, U.S. Patent No. 5,627,059, and U.S. Patent No. 5,631,153.

As will be appreciated by one of ordinary skill in the art, in certain circumstances it may be advantageous to modify an inventive polynucleotide sequence by ligating it to a heterologous sequence, thereby enabling the production of a fusion protein. Certain vectors are designed to incorporate such heterologous sequences so that insertion of a polynucleotide into the vector at an appropriate location results in an in-frame fusion to the heterologous sequence, which may be either upstream or downstream from the inserted polynucleotide. Such heterologous sequences may encode tags or cleavable linker sequences such as glutathione S-transferase (GST), the hemaglutinin epitope known as HA tag, a short stretch of the Myc protein (Myc tag), FLAG tag, 6X His tag, maltose binding protein tag, etc. In general, many of these tags are useful for the purification and/or detection of the polypeptide using an antibody or other reagent that binds to the tag. Other useful heterologous sequences include that of green fluorescent protein (GFP), which allows visualization of the fusion protein. The present invention encompasses all such BSTP-ECG1 fusion proteins.

The invention provides an antibody that binds to BSTP-ECG1 or to a fragment thereof. Antibodies to these polypeptides may be generated, for example, as described in Example 6 below. In general, such antibodies may be generated by methods well

known in the art and described, for example, in Harlow, E., and Lane, D., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1988. Details and references for the production of antibodies based on an inventive polypeptide may also be found in U.S. Patent No. 6,008,337. Antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric (e.g., "humanized"), and single chain antibodies, and Fab fragments. The invention encompasses "fully human" antibodies produced using the XenoMouse™ technology (AbGenix Corp., Fremont, CA) according to the techniques described in U.S. Patent No. 6,075,181.

B. Diagnostics and methods of use thereof

The invention provides reagents and methods for detecting the inventive polynucleotides and polypeptides described above. The reagents are useful for diagnostic purposes among others. In one aspect, the invention provides a method of classifying tumors by detecting the presence of one or more of the inventive polypeptides or polynucleotides, i.e., polypeptides or polynucleotides comprising a sequence set forth in SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 5, or fragments thereof. As is well known in the art, a polypeptide may be detected using a variety of techniques that employ an antibody that binds to the polypeptide. In addition, in certain embodiments of the invention the polypeptide is detected using other modalities directed at the detection of the polypeptide, such as aptamers (Aptamers, *Molecular Diagnosis*, Vol. 4, No. 4, 1999). In general, any appropriate method for detecting a polypeptide may be used in conjunction with the present invention, although antibodies represent a preferred modality.

Thus the invention provides a method for classifying a disease comprising the steps of: (a) obtaining cells or tissue from a site of disease; (b) detecting a polypeptide having a sequence selected from the group consisting of SEQ ID NO:1, or variants thereof; and (c) placing the disease into one of a set of predetermined categories based on detection of the polypeptide. The predetermined categories are, in general, characterized by differences in their expression of BSTP-ECG1 and also by differences in some aspect of tumor phenotype. For example, and without intending to be limiting, one predetermined category may be characterized by a good prognosis; one predetermined category may be characterized by a poor prognosis; and one

predetermined category may be characterized by a non-responder phenotype. In addition to differences in phenotype, the predetermined categories also feature differences in the expression of BSTP-ECG1, thus allowing the placement of a tumor into one of the categories based on the detection (which may include measurement) of BSTP-ECG1 in a sample from the tumor. Thus the assignment may be used as a basis for providing diagnostic, prognostic, and/or predictive information to a patient having the tumor.

The invention encompasses a number of uses for antibodies that bind to BSTP-ECG1. In a broad aspect, antibodies that bind to BSTP-ECG1 can be used to provide information useful for diagnosis, prognosis, classification, or monitoring of a disorder characterized by the inappropriate expression of the polypeptide. The term "inappropriate expression", as used herein, can include, but is not limited to, underexpression, overexpression, or expression in a cell type or organ in which the polypeptide is not normally expressed. Inappropriate expression can include (1) underexpression relative to normal for that cell or tissue, (2) overexpression relative to normal for that cell or tissue, or (3) mislocalization within a cell or tissue relative to normal for that cell or tissue. In order to provide a basis for diagnosis, prognosis, classification, or monitoring of a disorder characterized by the inappropriate expression of the BSTP-ECG1 polypeptide, a normal or standard expression profile can first be established by measuring the expression of BSTP-ECG1 in cells, tissues, body fluids, etc., obtained from subjects not suffering from the disorder. In general, a range of values may be considered normal, and departure from within this range of values may be taken to indicate that an individual suffers from or is at increased likelihood to develop the disorder. Of course in certain cases the information will simply consist of an indication of the presence or absence of BSTP-ECG1 in a cell or tissue sample (e.g., a sample of breast cancer cells or tissue), regardless of whether BSTP-ECG1 expression is considered "inappropriate".

As used herein the term "diagnostic information" includes, but is not limited to, any type of information that is useful in determining whether a patient has, or is at increased risk for developing, a disease or disorder; for providing a prognosis for a patient having a disease or disorder; for classifying a disease or disorder; for monitoring a patient for recurrence of a disease or disorder; for selecting a preferred

therapy; for predicting the likelihood of response to a therapy, etc. In a preferred embodiment of the invention, antibodies to BSTP-ECG1 are used for providing diagnostic information for cancer, particularly for breast cancer. In general, diagnostic assays in which the antibodies may be employed include methods that use the antibody to detect BSTP-ECG1 in a tissue sample, cell sample, body fluid sample (e.g., serum), cell extract, etc. Thus the invention provides a method for detecting a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1 in a biological sample comprising steps of: (a) contacting the biological sample with an antibody that binds to the polypeptide of SEQ ID NO:1; and (b) determining whether the antibody specifically binds to the sample, the binding being an indication that the sample contains the polypeptide. The biological sample can be processed in any of a variety of ways prior to being placed in contact with the antibody.

Many detection methods typically involve the use of a labeled secondary antibody that recognizes the primary antibody (i.e., the antibody that binds to the polypeptide being detected). Depending upon the nature of the sample, appropriate methods include, but are not limited to, immunohistochemistry, radioimmunoassay, ELISA, immunoblotting, and FACS analysis. In the case where the polypeptide is to be detected in a tissue sample, e.g., a biopsy sample, immunohistochemistry is a particularly appropriate detection method. Techniques for obtaining tissue and cell samples and performing immunohistochemistry and FACS are well known in the art. Such techniques are routinely used, for example, to detect the ER in breast tumor tissue or cell samples. In general, such tests will include a negative control, which can involve applying the test to normal tissue so that the signal obtained thereby can be compared with the signal obtained from the sample being tested. In tests in which a secondary antibody is used to detect the antibody that binds to the polypeptide of interest, an appropriate negative control can involve performing the test on a portion of the sample with the omission of the antibody that binds to the polypeptide to be detected, i.e., with the omission of the primary antibody. Antibodies suitable for use as diagnostics generally exhibit high specificity for the target polypeptide and low background. In general, monoclonal antibodies are preferred for diagnostic purposes.

In general, the results of such a test can be presented in any of a variety of formats. The results can be presented in a qualitative fashion. For example, the test report may indicate only whether or not a particular polypeptide was detected, perhaps also with an indication of the limits of detection. The results may be presented in a semi-quantitative fashion. For example, various ranges may be defined, and the ranges may be assigned a score (e.g., 1+ to 4+) that provides a certain degree of quantitative information. Such a score may reflect various factors, e.g., the number of cells in which the polypeptide is detected, the intensity of the signal (which may indicate the level of expression of the polypeptide), etc. The results may be presented in a quantitative fashion, e.g., as a percentage of cells in which the polypeptide is detected, as a protein concentration, etc. As will be appreciated by one of ordinary skill in the art, the type of output provided by a test will vary depending upon the technical limitations of the test and the biological significance associated with detection of the polypeptide. For example, in the case of certain polypeptides a purely qualitative output (e.g., whether or not the polypeptide is detected at a certain detection level) provides significant information. In other cases a more quantitative output (e.g., a ratio of the level of expression of the polypeptide in the sample being tested versus the normal level) is necessary.

A particular use for antibodies that bind to BSTP-ECG1 is to classify breast tumors based on the association between the expression level of the *BST-ECG1* gene, or the association between a gene subset that includes the *BST-ECG1* gene, and a tumor subset having a particular phenotype (e.g., a good prognosis phenotype, a poor prognosis phenotype, a non-responder phenotype, etc.). Immunohistochemistry using an antibody that binds to BSTP-ECG1 can be employed to detect BSTP-ECG1 in a tumor tissue or cell sample, thereby providing information useful in determining whether the tumor expresses or overexpresses the polypeptide. Additional detection methods including RIA, ELISA, immunoblotting, and FACS analysis. Thus the present invention provides a test for classifying tumors.

The result of such a test (e.g., whether a given tumor expresses or overexpresses BSTP-ECG1, quantitative expression level of BSTP-ECG1, etc.) can be used to provide information about the prognosis of the tumor or the likelihood that the tumor will respond to therapy. In certain embodiments of the inventive methods a

single antibody is used whereas in other embodiments of the invention multiple antibodies, directed either against the same or against different polypeptides can be used to increase the sensitivity or specificity of the test or to provide more detailed information than that provided by a single antibody. Thus the invention encompasses the use of a battery of antibodies, one or more of which binds to BSTP-ECG1.

Although in many cases detection of polypeptides using antibodies represents the most convenient means of determining whether *BST-ECG1* is expressed (or overexpressed) in a particular sample, the invention also encompasses the use of polynucleotides for this purpose. The invention provides methods for detecting a polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a fragment thereof, in a biological sample. One such method comprises steps of: (a) hybridizing a nucleic acid complementary to the polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a fragment thereof, to at least one nucleic acid in the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex indicates the presence of a polynucleotide encoding the polypeptide in the biological sample. A second such method comprises steps of:

(a) hybridizing a nucleic acid encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or a fragment thereof, to at least one nucleic acid complementary to a nucleic acid in the biological sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex indicates the presence of a polynucleotide encoding the polypeptide in the biological sample. In other words, detection can comprise detecting either a polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, or detecting the complement of such a polynucleotide. For example, in certain embodiments of the inventive method mRNA in the biological sample is detected while in other embodiments of the invention cDNA synthesized from mRNA in the biological sample is detected.

The hybridization complex can be detected in any of a variety of ways. For example, the hybridization complex may be formed on a microarray (e.g., a cDNA

array or an oligonucleotide array) and detected using techniques such as those described herein or related techniques well known in the art. Microarray analysis is but one means by which polynucleotides can be used to detect or measure *BST-ECG1* expression. Expression of *BST-ECG1* can also be measured by a variety of other techniques that make use of a polynucleotide corresponding to part or all of the *BST-ECG1* gene rather than an antibody that binds to a polypeptide encoded by the gene. Appropriate techniques include, but are not limited to, *in situ* hybridization, Northern blot, and various nucleic acid amplification techniques such as PCR, quantitative PCR, and the ligase chain reaction.

Another aspect of the invention comprises a kit to test for the presence of any of the inventive polynucleotides or polypeptides, e.g., in a tissue sample or in a body fluid. The kit can comprise, for example, an antibody for detection of a polypeptide (e.g., the polypeptide of SEQ ID NO:1) or a probe for detection of a polynucleotide. In addition, the kit can comprise a reference sample, instructions for processing samples, performing the test and interpreting the results, buffers and other reagents necessary for performing the test. In certain embodiments the kit can comprise a panel of antibodies. In certain embodiments the kit comprises a cDNA or oligonucleotide array for detection of expression of the gene encoding BSTP-ECG1, e.g., for detecting the presence of a polynucleotide comprising the nucleotide sequence set forth in SEQ ID NO:2, SEQ ID NO:3, or SEQ ID NO: 4.

Yet another aspect of the invention comprises detecting mutations in the gene encoding BSTP-ECG1 and/or in regulatory regions of the gene. The invention further encompasses detecting allelic variants of the gene encoding BSTP-ECG1. As mentioned above, mutations in certain genes (e.g., *BRCA-1*, *BRCA-2*) have been associated with an increased risk of breast cancer. The detection of mutations and allelic variants can be performed using any of a variety of methods well known in the art ranging from use of microarrays (e.g., oligonucleotide arrays) to detect single nucleotide polymorphisms (SNPs) associated with a particular allele, use of microarrays (e.g.,oligonucleotide arrays) to detect substitutions, deletions, etc., detection of restriction fragment length polymorphisms (RFLPs), direct sequencing of DNA isolated from an individual, etc.

C. Therapeutics

The invention includes the use of the polynucleotides, polypeptides, and antibodies described herein as therapeutic agents for the treatment of cancer. In particular, the invention contemplates the use of the polynucleotides, polypeptides, and antibodies described herein for the treatment of breast cancer, although their use for the treatment of other forms of cancer is also within the scope of the invention. The invention specifically encompasses antagonists to BSTP-ECG1. Such antagonists (which include, but are not limited to, antibodies, small molecules, antisense nucleic acids) may be produced or identified using any of a variety of methods known in the art. For example, a purified inventive BSTP-ECG1 polypeptide or fragment thereof may be used to raise antibodies or to screen libraries of compounds to identify those that specifically bind to the polypeptide. While not wishing to be bound by any theory, the presence of a putative transmembrane domain suggesting that BSTP-ECG1 is likely to include an extracellular region may make it a particularly preferred target for therapeutics.

Preferably antibodies suitable for use as therapeutics exhibit high specificity for the target polypeptide and low background binding to other polypeptides. In general, monoclonal antibodies are preferred for therapeutic purposes. In the case of breast cancer, antibodies against the HER2/neu/ErbB2 polypeptide (a polypeptide homologous to the epidermal growth factor receptor) represent a paradigm in terms of the development of therapeutic antibodies. The *HER2/neu/ErbB2* gene is overexpressed in approximately 25 to 30 percent of metastatic breast tumors, and an antibody against the HER2/neu/ErbB2 polypeptide, Herceptin® (Trastuzumab) is approved for the treatment of certain patients with metastatic breast cancer, confirming the utility of therapeutic antibodies directed against polypeptides that are specifically overexpressed in particular tumors subsets. Antibodies directed against a polypeptide expressed by a cell may have a number of mechanisms of action. In certain instances, e.g., in the case of a polypeptide that exerts a growth stimulatory effect on a cell, antibodies may directly antagonize the effect of the polypeptide and thereby arrest tumor progression, trigger apoptosis, etc. While not wishing to be bound by any theory, it may be particularly likely that certain genes that are overexpressed in tumors having a poor prognosis encode polypeptides that have a

growth stimulatory effect on tumor cells or facilitate the growth of such cells in some other way, e.g., by enhancing angiogenesis, by allowing cells to overcome normal growth regulatory mechanisms, or by blocking mechanisms that would normally lead to elimination of mutated or otherwise abnormal cells. In certain embodiments of the invention the antibody may serve to target a toxic moiety to the cell. Thus the invention encompasses the use of antibodies that have been conjugated with a cytotoxic agent, e.g., a toxin such as ricin or diphtheria toxin, a radioactive moiety, etc. Such antibodies can be used to direct the cytotoxic agent specifically to cells that express the inventive polypeptide.

Although certain antagonists may function through direct interaction with a polypeptide such as BSTP-ECG1, e.g., by inhibiting its activity, others may function by affecting expression of the polypeptide. Reduction in expression of an endogenously produced polypeptide may be achieved by the administration of antisense nucleic acids (e.g., oligonucleotides, RNA, DNA, most typically oligonucleotides that have been modified to improve stability or targeting) or peptide nucleic acids comprising sequences complementary to those of the mRNA that encodes the polypeptide. Antisense technology and its applications are described in Phillips, M.I. (ed.) *Antisense Technology*, Methods Enzymol., Volumes 313 and 314, Academic Press, San Diego, 2000, and references mentioned therein. Ribozymes (catalytic RNA molecules that are capable of cleaving other RNA molecules) represent another approach to reducing gene expression. Such ribozymes can be designed to cleave specific mRNAs corresponding to a gene of interest. Their use is described in U.S. Patent No. 5,972,621, and references therein. The invention encompasses the delivery of antisense and/or ribozyme molecules via a gene therapy approach in which vectors or cells expressing the antisense molecules are administered to an individual.

It may also be desirable to increase the expression of a gene encoding BSTP-ECG1 or to increase the activity of BSTP-ECG1. For example, in the case of genes (such as that encoding BSTP-ECG1), whose expression correlates with that of the estrogen receptor and therefore is associated with a good prognosis, it may be desirable to increase the expression of such genes or the activity of the corresponding polypeptides in tumors that fail to express these genes.

Small molecule modulators (e.g., inhibitors or activators) of *BST-ECG1* gene expression are also within the scope of the invention and may be detected by screening libraries of compounds using, for example, cell lines that express the polypeptide or a version of the polypeptide that has been modified to include a readily detectable moiety. Methods for identifying compounds capable of modulating gene expression are described, for example, in U.S. Patent No. 5,976,793. The screening methods described therein are particularly appropriate for identifying compounds that do not naturally occur within cells and that modulate the expression of genes of interest whose expression is associated with a defined physiological or pathological effect within a multicellular organism.

More generally, the invention encompasses compounds that modulate the activity of a polypeptide encoding BSTP-ECG1. Methods of screening for such interacting compounds are well known in the art and depend, to a certain degree, on the particular properties and activities of the polypeptide encoded by the gene. Representative examples of such screening methods may be found, for example, in U.S. Patent No. 5,985,829, U.S. Patent No. 5,726,025, U.S. Patent No. 5,972,621, and U.S. Patent No. 6,015,692. The skilled practitioner will readily be able to modify and adapt these methods as appropriate for BSTP-ECG1. The mechanism of modulation need not be direct. For example, the modulator may act on an enzyme that may modify BSTP-ECG1.

The invention also encompasses the use of polynucleotides encoding BSTP-ECG1, or portions thereof, as DNA vaccines. Such vaccines comprise polynucleotide sequences, typically inserted into vectors, that direct the expression of an antigenic polypeptide within the body of the individual being immunized. Details regarding the development of vaccines, including DNA vaccines, for various forms of cancer may be found, for example, in Brinckerhoff L.H., Thompson L.W., Slingluff C.L., Jr., Melanoma Vaccines, *Curr Opin Oncol*, 12(2):163-73, 2000 and in Stevenson, F.K., DNA vaccines against cancer: from genes to therapy, *Ann. Oncol.*, 10(12): 1413-8, 1999 and references therein. BSTP-ECG1 polypeptides, or fragments thereof, that may also find use as cancer vaccines. Any of these vaccines may be used for the prevention and/or the treatment of cancer.

The invention includes pharmaceutical compositions comprising the inventive polypeptides, polynucleotides, antibodies, small molecule inhibitors, agonists, or antagonists described above. In general, a pharmaceutical composition will include an active agent in addition to one or more inactive agents such as a sterile, biocompatible carrier including, but not limited to, sterile water, saline, buffered saline, or dextrose solution. The pharmaceutical compositions may be administered either alone or in combination with other therapeutic agents including other chemotherapeutic agents, hormones, vaccines, and/or radiation therapy. By "in combination with", it is not intended to imply that the agents must be administered at the same time or formulated for delivery together, although these methods of delivery are within the scope of the invention. In general, each agent will be administered at a dose and on a time schedule determined for that agent. Additionally, the invention encompasses the delivery of the inventive pharmaceutical compositions in combination with agents that may improve their bioavailability, reduce or modify their metabolism, inhibit their excretion, or modify their distribution within the body. The invention encompasses treating cancer, particularly breast cancer, by administering the pharmaceutical compositions of the invention. Although the pharmaceutical compositions of the present invention can be used for treatment of any subject (e.g., any animal) in need thereof, they are most preferably used in the treatment of humans.

The pharmaceutical compositions of this invention can be administered to humans and other animals by a variety of routes including oral, intravenous, intramuscular, intraarterial, subcutaneous, intraventricular, transdermal, rectal intravaginal, intraperitoneal, topical (as by powders, ointments, or drops), bucal, or as an oral or nasal spray or aerosol. In general the most appropriate route of administration will depend upon a variety of factors including the nature of the compound (e.g., its stability in the environment of the gastrointestinal tract), the condition of the patient (e.g., whether the patient is able to tolerate oral administration), etc. At present the intravenous route is most commonly used to deliver therapeutic antibodies and nucleic acids. However, the invention encompasses the delivery of the inventive pharmaceutical composition by any appropriate route taking into consideration likely advances in the sciences of drug delivery.

General considerations in the formulation and manufacture of pharmaceutical agents may be found, for example, in *Remington's Pharmaceutical Sciences*, 19th ed., Mack Publishing Co., Easton, PA, 1995. It will be appreciated that certain of the compounds of the present invention can exist in free form for treatment, or, where appropriate, in salt form, as discussed in more detail below. Compounds to be utilized in the pharmaceutical compositions include compounds existing in free form or pharmaceutically acceptable derivatives thereof, as defined herein, such as pharmaceutically acceptable salts, esters, salts of such esters, or any other adduct or derivative, which upon administration to a patient in need, is capable of providing, directly or indirectly, a compound as otherwise described herein, or a metabolite or residue thereof, e.g., a prodrug. Thus, as used herein, the term "pharmaceutically acceptable salt" refers to those salts which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of humans and lower animals without undue toxicity, irritation, allergic response and the like, and are commensurate with a reasonable benefit/risk ratio. Pharmaceutically acceptable salts are well known in the art. For example, S. M. Berge, et al. describe pharmaceutically acceptable salts in detail in *J. Pharmaceutical Sciences*, 66: 1-19 (1977), incorporated herein by reference. The salts can be prepared in situ during the final isolation and purification of the compounds of the invention, or separately by reacting the free base function with a suitable organic acid. Examples of pharmaceutically acceptable, nontoxic acid addition salts are salts of an amino group formed with inorganic acids such as hydrochloric acid, hydrobromic acid, phosphoric acid, sulfuric acid and perchloric acid or with organic acids such as acetic acid, oxalic acid, maleic acid, tartaric acid, citric acid, succinic acid, or malonic acid or by using other methods used in the art such as ion exchange. Other pharmaceutically acceptable salts include adipate, alginate, ascorbate, aspartate, benzenesulfonate, benzoate, bisulfate, borate, butyrate, camphorate, camphorsulfonate, citrate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, formate, fumarate, glucoheptonate, glycerophosphate, gluconate, hemisulfate, heptanoate, hexanoate, hydroiodide, 2-hydroxy-ethanesulfonate, lactobionate, lactate, laurate, lauryl sulfate, malate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oleate, oxalate, palmitate, pamoate, pectinate, persulfate, 3-phenylpropionate, phosphate,

picrate, pivalate, propionate, stearate, succinate, sulfate, tartrate, thiocyanate, p-toluenesulfonate, undecanoate, valerate salts, and the like. Representative alkali or alkaline earth metal salts include sodium, lithium, potassium, calcium, magnesium, and the like. Further pharmaceutically acceptable salts include, when appropriate, nontoxic ammonium, quaternary ammonium, and amine cations formed using counterions such as halide, hydroxide, carboxylate, sulfate, phosphate, nitrate, lower alkyl sulfonate and aryl sulfonate.

Additionally, as used herein, the term "pharmaceutically acceptable ester" refers to esters that hydrolyze in vivo and include those that break down readily in the human body to leave the parent compound or a salt thereof. Suitable ester groups include, for example, those derived from pharmaceutically acceptable aliphatic carboxylic acids, particularly alkanoic, alkenoic, cycloalkanoic and alkanedioic acids, in which each alkyl or alkenyl moiety advantageously has not more than 6 carbon atoms. Examples of particular suitable esters includes formates, acetates, propionates, butyrates, acrylates and ethylsuccinates.

Furthermore, the term "pharmaceutically acceptable prodrugs" as used herein refers to those prodrugs of the compounds of the present invention that are, within the scope of sound medical judgment, suitable for use in contact with the tissues of humans and lower animals without undue toxicity, irritation, allergic response, and the like, commensurate with a reasonable benefit/risk ratio, and effective for their intended use, as well as the zwitterionic forms, where possible, of the compounds of the invention. The term "prodrug" refers to compounds that are rapidly transformed *in vivo* to yield a particular active compound, for example by hydrolysis in blood. A thorough discussion is provided in T. Higuchi and V. Stella, "Pro-drugs as Novel Delivery Systems", Vol. 14 of the A.C.S. Symposium Series, and in Edward B. Roche, ed., *Bioreversible Carriers in Drug Design*, American Pharmaceutical Association and Pergamon Press, 1987, both of which are incorporated herein by reference.

As mentioned above, the pharmaceutical compositions of the present invention additionally comprise a pharmaceutically acceptable carrier, which, as used herein, means a non-toxic, inert solid, semi-solid or liquid filler, diluent, encapsulating material, or formulation auxiliary of any type. Some examples of materials which can

serve as pharmaceutically acceptable carriers are sugars such as lactose, glucose and sucrose; starches such as corn starch and potato starch; cellulose and its derivatives such as sodium carboxymethyl cellulose, ethyl cellulose and cellulose acetate; powdered tragacanth; malt; gelatin; talc; excipients such as cocoa butter and suppository waxes; oils such as peanut oil, cottonseed oil; safflower oil; sesame oil; olive oil; corn oil and soybean oil; glycols; such a propylene glycol; esters such as ethyl oleate and ethyl laurate; agar; buffering agents such as magnesium hydroxide and aluminum hydroxide; alginic acid; water; isotonic saline; Ringer's solution; ethyl alcohol, and phosphate buffer solutions, dextrose solutions, as well as other non-toxic compatible lubricants such as sodium lauryl sulfate and magnesium stearate, as well as coloring agents, releasing agents, coating agents, sweetening, flavoring and perfuming agents, preservatives and antioxidants can also be present in the composition, according to the judgment of the formulator.

Liquid dosage forms for oral administration include pharmaceutically acceptable emulsions, microemulsions, solutions, suspensions, syrups and elixirs. In addition to the active compounds, the liquid dosage forms may contain inert diluents commonly used in the art such as, for example, water or other solvents, solubilizing agents and emulsifiers such as ethyl alcohol, isopropyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils (in particular, cottonseed, groundnut, corn, germ, olive, castor, and sesame oils), glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols and fatty acid esters of sorbitan, and mixtures thereof. Besides inert diluents, the oral compositions can also include adjuvants such as wetting agents, emulsifying and suspending agents, sweetening, flavoring, and perfuming agents.

Injectable preparations, for example, sterile injectable aqueous or oleaginous suspensions may be formulated according to the known art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile injectable solution, suspension or emulsion in a nontoxic parenterally acceptable diluent or solvent, for example, as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, U.S.P. and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any

bland fixed oil can be employed including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid are used in the preparation of injectables.

The injectable formulations can be sterilized, for example, by filtration through a bacterial-retaining filter, or by incorporating sterilizing agents in the form of sterile solid compositions which can be dissolved or dispersed in sterile water or other sterile injectable medium prior to use.

In order to prolong the effect of a drug, it is often desirable to slow the absorption of the drug from subcutaneous or intramuscular injection. This may be accomplished by the use of a liquid suspension of crystalline or amorphous material with poor water solubility. The rate of absorption of the drug then depends upon its rate of dissolution which, in turn, may depend upon crystal size and crystalline form. Alternatively, delayed absorption of a parenterally administered drug form is accomplished by dissolving or suspending the drug in an oil vehicle. Injectable depot forms are made by forming microencapsulated matrices of the drug in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in liposomes or microemulsions which are compatible with body tissues.

Compositions for rectal or vaginal administration are preferably suppositories which can be prepared by mixing the compounds of this invention with suitable non-irritating excipients or carriers such as cocoa butter, polyethylene glycol or a suppository wax which are solid at ambient temperature but liquid at body temperature and therefore melt in the rectum or vaginal cavity and release the active compound.

Solid dosage forms for oral administration include capsules, tablets, pills, powders, and granules. In such solid dosage forms, the active compound is mixed with at least one inert, pharmaceutically acceptable excipient or carrier such as sodium citrate or dicalcium phosphate and/or a) fillers or extenders such as starches, lactose, sucrose, glucose, mannitol, and silicic acid, b) binders such as, for example, carboxymethylcellulose, alginates, gelatin, polyvinylpyrrolidinone, sucrose, and

acacia, c) humectants such as glycerol, d) disintegrating agents such as agar–agar, calcium carbonate, potato or tapioca starch, alginic acid, certain silicates, and sodium carbonate, e) solution retarding agents such as paraffin, f) absorption accelerators such as quaternary ammonium compounds, g) wetting agents such as, for example, cetyl alcohol and glycerol monostearate, h) absorbents such as kaolin and bentonite clay, and i) lubricants such as talc, calcium stearate, magnesium stearate, solid polyethylene glycols, sodium lauryl sulfate, and mixtures thereof. In the case of capsules, tablets and pills, the dosage form may also comprise buffering agents.

Solid compositions of a similar type may also be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polyethylene glycols and the like. The solid dosage forms of tablets, dragees, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings and other coatings well known in the pharmaceutical formulating art. They may optionally contain opacifying agents and can also be of a composition that they release the active ingredient(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed manner. Examples of embedding compositions that can be used include polymeric substances and waxes. Solid compositions of a similar type may also be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polethylene glycols and the like.

The active compounds can also be in micro-encapsulated form with one or more excipients as noted above. The solid dosage forms of tablets, dragees, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings, release controlling coatings, and other coatings well known in the pharmaceutical formulating art. In such solid dosage forms the active compound may be admixed with at least one inert diluent such as sucrose, lactose or starch. Such dosage forms may also comprise, as is normal practice, additional substances other than inert diluents, e.g., tableting lubricants and other tableting aids such a magnesium stearate and microcrystalline cellulose. In the case of capsules, tablets and pills, the dosage forms may also comprise buffering agents. They may optionally contain opacifying agents and can also be of a composition that they release the active ingredient(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed

manner. Examples of embedding compositions that can be used include polymeric substances and waxes.

Dosage forms for topical or transdermal administration of a compound of this invention include ointments, pastes, creams, lotions, gels, powders, solutions, sprays, inhalants or patches. The active component is admixed under sterile conditions with a pharmaceutically acceptable carrier and any needed preservatives or buffers as may be required. Ophthalmic formulation and ear drops are also contemplated as being within the scope of this invention. The ointments, pastes, creams and gels may contain, in addition to an active compound of this invention, excipients such as animal and vegetable fats, oils, waxes, paraffins, starch, tragacanth, cellulose derivatives, polyethylene glycols, silicones, bentonites, silicic acid, talc and zinc oxide, or mixtures thereof. Powders and sprays can contain, in addition to the compounds of this invention, excipients such as lactose, talc, silicic acid, aluminum hydroxide, calcium silicates and polyamide powder, or mixtures of these substances. Sprays can additionally contain propellants known in the art such as chlorofluorohydrocarbons.

Transdermal patches have the added advantage of providing controlled delivery of a compound to the body. Such dosage forms can be made by dissolving or dispensing the compound in the proper medium. Absorption enhancers can also be used to increase the flux of the compound across the skin. The rate can be controlled by either providing a rate controlling membrane or by dispersing the compound in a polymer matrix or gel.

In yet another aspect, the present invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention, and in certain embodiments, includes an additional approved therapeutic agent for use as a combination therapy. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceutical products, which notice reflects approval by the agency of manufacture, use or sale for human administration. Instructions for use of the compound(s) may also be included.

According to the methods of treatment of the present invention, cancer, particularly breast cancer, is treated or prevented in a patient such as a human or other

mammal by administering to the patient a therapeutically effective amount of a compound of the invention, in such amounts and for such time as is necessary to achieve the desired result. By a "therapeutically effective amount" of a compound of the invention is meant a sufficient amount of the compound to treat (e.g. to ameliorate the symptoms of, delay progression of, prevent recurrence of, cure, etc.) cancer, particularly breast cancer, at a reasonable benefit/risk ratio, which involves a balancing of the efficacy and toxicity of the compound. In general, therapeutic efficacy and toxicity may be determined by standard pharmacological procedures in cell cultures or with experimental animals, e.g., by calculating the ED_{50} (the dose that is therapeutically effective in 50% of the treated subjects) and the LD_{50} (the dose that is lethal to 50% of treated subjects). The ED_{50}/LD_{50} represents the therapeutic index of the compound. Although in general drugs having a large therapeutic index are preferred, as is well known in the art, a smaller therapeutic index may be acceptable in the case of a serious disease, particularly in the absence of alternative therapeutic options. Ultimate selection of an appropriate range of doses for administration to humans is determined in the course of clinical trials.

It will be understood that the total daily usage of the compounds and compositions of the present invention for any given patient will be decided by the attending physician within the scope of sound medical judgment. The specific therapeutically effective dose level for any particular patient will depend upon a variety of factors including the disorder being treated and the severity of the disorder; the activity of the specific compound employed; the specific composition employed; the age, body weight, general health, sex and diet of the patient; the time of administration, route of administration, and rate of excretion of the specific compound employed; the duration of the treatment; drugs used in combination or coincidental with the specific compound employed; and like factors well known in the medical arts.

The total daily dose of the compounds of this invention administered to a human or other mammal in single or in divided doses can be in amounts, for example, from 0.01 to 50 mg/kg body weight or more usually from 0.1 to 25 mg/kg body weight. Single dose compositions may contain such amounts or submultiples thereof to make up the daily dose. In general, treatment regimens according to the present

invention comprise administration to a patient in need of such treatment from about 0.1 μg to about 2000 mg of the compound(s) of the invention per day in single or multiple doses.

5

EXAMPLES

Note: A numbered list of references for the Examples appears following the Examples, all of which are incorporated herein by reference.

Example 1

Preparation of Microarrays Containing 8498 Human cDNAs

The human cDNA clones used in this study were obtained from Research Genetics (Huntsville AB, USA) as bacterial colonies in 96-well microtiter plates. The clones were chosen from a set of 15,000 cDNA clones that corresponded to the Research Genetics Human Gene Filters sets GF200-202 (http://www.resgen.com/). These clones form part of a set of clones assembled by the I.M.A.G.E. consortium (Lennon, G.G., Auffray, C., Polymeropoulos, M., Soares, M.B. The I.M.A.G.E. Consortium: An Integrated Molecular Analysis of Genomes and their Expression. Genomics 33:151-152,1996) and are identified by I.M.A.G.E. clone ID numbers. All clones printed on these arrays were sequence validated as part of a product offered at Research Genetics, Inc. We estimate that greater than 97% of the clones on the array are correctly identified.

A detailed protocol for the production of the cDNA microarrays used in this study is available at http://cmgm.stanford.edu/pbrown/protocols.html and is reproduced below with insubstantial changes. As described below, the protocol includes steps of (1) cleaning the glass slides onto which the DNAs (e.g., products of PCR reactions) are to be spotted; (2) spotting the DNAs onto the glass slides with an arrayer; (3) Post processing to prepare arrays containing spotted DNAs for hybridization. All procedures are done at room temperature and with double distilled water unless otherwise stated. Unless otherwise stated, in this Example and the following Examples, reagents are prepared according to protocols available in Maniatis, T., Sambrook, J. and Fritsch, E., *Molecular Cloning: A Laboratory Manual* (3 Volume Set), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989, the contents of which are herein incorporated by reference.

Cleaning Slides

Use 30 slide racks in 350mL glass dishes

1. Dissolve 50g of NaOH pellets into 150ml ddH2O
2. Add 200ml of 95% EtOH, stir until completely mixed
3. If solution remains cloudy, add ddH2O until clear
4. Pour solution into glass slide box.
5. Drop in 30 slides in a metal rack. (Gold Seal slides, Cat. 3010)
6. Let soak on an orbital shaker for at least two hours
7. Rinse slides by transferring rack to slide dish filled with ddH2O
8. Repeat ddH2O rinses x3. It's important to remove all traces of the NaOH-ethanol.
9. Prepare Poly-l-lysine solution: Use Sigma Poly-l-lysine solution. Cat. No. 8920
10. Add 70mL poly-l-lysine to 280ml of water
11. Transfer slides to poly-l-lysine solution and let soak for 1 hour.
12. Remove excess liquid from slides by spinning the rack of slides on microtiter plate carriers at 500rpm.
13. Dry slides at 40 degrees C for 5 minutes in a vacuum oven.
14. Store slides in a closed box for at least two weeks prior to use.
15. Before printing arrays, check a sample slide to make sure it's hydrophobic (water should bead off it) but the lysine coating is not turning opaque.

Arraying

1. Transfer PCR reactions to 96-well V-bottom tissue culture plates (Costar). Add 1/10 vol. 3M sodium acetate (pH 5.2) and equal volume isopropanol. Store at -20 C for a few hours.
2. Centrifuge in Sorvall at 3500 RPM for 45 min. Rinse with 70% EtOH, centrifuge again and dry.
2. Resuspend DNA in 12ul 3X SSC for a few hours and transfer to flexible U-bottom printing plates.
4. Spot DNA onto poly-l-lysine slides with an arrayer.

Post processing

1. Rehydrate arrays by suspending slides over a dish of warm ddH2O. (~1 minute)

2. Snap-dry each array (DNA side up) on a 100C hot plate for 3 seconds.

3. UV cross-link DNA to the glass by using a Stratalinker set for 60 milliJoules

4. Dissolve 5g of succinic anhydride (Aldrich) in 315mL of n-methyl-pyrrolidinone.

5. To this, add 35mL of 0.2M NaBorate pH 8.0 (made by dissolving boric acid in water and adjusting the pH with NaOH), and stir until dissolved.

6. Soak arrays in this solution for 15 minutes with shaking.

7. Transfer arrays to 95C water bath for 2 minutes

8. Quickly transfer arrays to 95% EtOH for 1 minute.

9. Remove excess liquid from slides by spinning the rack of slides on microtiter plate carriers at 500rpm.

10. Arrays can be used immediately.

Reagent Suppliers

Microscope slides Goldseal brand. (Cat. 3010)
Poly-l-lysine solution Sigma product number P8920
Succinic Anhydride Aldrich product number 23,969-0
N-Methyl-Pyrrolidinone Aldrich product number 32,863-4

Microarrays were prepared according to the above protocol using the 8498 cDNA clones described above. All microarrays used in the experiments described herein were from a single print run batch of microarrays.

Example 2

Cell Lines, Breast Tissue, and Breast Tumor Samples for Microarray Analysis and Preparation of mRNA Samples

Common Reference Sample

Each of the 84 experimental samples tested here was analyzed by a comparative hybridization, using a common reference RNA pool as a standard; this reference sample was composed of equal mixtures of mRNA isolated from 11 established cell lines derived from human tissue (MCF7, Hs578T, OVCAR3, HepG2, NTERA2, MOLT4, RPMI-8226, NB4+ATRA, UACC-62, SW872, and Colo205: also see Table 2 for more details). The 11 cell lines were all grown to 70-90% confluence in RPMI medium, containing 10% Fetal Calf Serum and Penicillin/Streptomycin. The cells were harvested either by scraping or centrifugation, quickly resuspended in RNA lysis buffer and mRNA prepared using the FastTrack™ 2.0 mRNA Isolation Kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. In each case, multiple individual mRNA preparations were collected for each cell line, which were then pooled together and analyzed via Northern analysis before final mixing to ensure the quality of the input mRNAs (e.g., to confirm that the mRNA exhibited a size distribution indicating that it was substantially nondegraded). The 11 mRNA samples were then mixed together in equal amounts, aliquoted in 10mM Tris (7.4), and stored at -80 C until use (2 micrograms of common reference sample was used per microarray hybridization and was always labeled using Cy3).

Normal Breast Tissue

Three samples of normal breast tissue were analyzed. Two of the samples were obtained from Clontech (Palo Alto, CA) and were pools of six (Normal1) or two (Normal2) whole normal breasts. The third sample (Normal3) was obtained from a single individual.

Breast Tumor Samples

The 40 individual breast tumor samples were collected at either Stanford University in Stanford CA, USA, or in the Haukeland University Hospital in Bergen, Norway. Twenty of the forty breast tumors were sampled twice as part of a larger

Norwegian study on locally advanced breast cancers (T3/T4 and/or N2 tumors) and have been described previously (Aas, T., et al., *Nat. Med.*, 2, 811-814, 1996, the contents of which are incorporated herein by reference) ; these patients underwent an open surgical biopsy before treatment with doxorubicin monotherapy (range 12-23 weeks), followed by the definitive surgical resection of the remaining tumor after therapy, and were evaluated for clinical responses according to UICC criteria (Hayward, J., et al., *Br. J. Cancer*, 35, 292-298, 1977). In addition to the 20 pairs, there were 8 additional "before" specimens from Norway and 10 tumor specimens from Stanford (all Stanford tumors tested had a diameter of 3 cm or larger). Finally, 2 of the 10 Stanford tumor specimens assayed were also paired with a lymph node metastasis from the same patient.

mRNA Isolation from Breast Tumor and Tissue Samples

Following their excision, breast tumor samples were rapidly frozen in liquid N2 and then stored at -80 C until use. mRNA was isolated from breast tumors and normal breast tissue using the Trizol Reagent (Gibco-BRL) and Invitrogen FastTrack 2.0 Kit (all Stanford samples, and see http://genome-www.stanford.edu/sbcmp/web.shtml for the detailed protocol) or using the Trizol Reagent followed by Dynal bead separation for the mRNA purification step (all Norway tissue samples). Briefly, frozen tumor samples were cut into small pieces and immediately placed into 12 ml of Trizol Reagent. Each tumor sample in Trizol was homogenized using a PowerGen 125 Tissue Homogenizer (Fisher Scientific), and total RNA was isolated according to the Trizol reagent manufacturer's protocol. Tumor mRNA was isolated according to the manufacturer's protocols using the FastTrack 2.0 Kit (Invitrogen) or Dynal beads.

Example 3

Characterization of Breast Tissue and Tumor Samples

For all but two of the tumor specimens (i.e. New York 1 and New York 2), the mutational status of the *TP53* gene was determined using published methods (Aas, T., *et al.*).

A single pathologist (applicant Matt van de Rijn) reviewed hematoxylin and eosin (H&E) sections of each tumor, including all before and after pairs, and made a histological evaluation of each while blinded to the source. Tumors were graded using a modified version of the Bloom-Richardson method (Robbins, P., et al., *Hum Pathol*, 26, 873-879, 1995). These data are displayed in Table 4. Representative H&E sections of each tumor are posted on Applicants' website at http://genome-www.stanford.edu/molecularportraits/.

Immunohistochemistry was performed as described previously (Perou, C., *et al.*, 1999; Bindl, J. and Warnke, R., *Am J Clin Pathol*, 85, 490-493, 1986, and Natkunam, Y., *et al.*, Am. J. Path., 156(1), 2000, the contents of which are incorporated herein by reference). The antibodies used included the commercially available monoclonal antibodies CAM5.2 (specific for keratins 8/18, available from Becton Dickinson), anti-keratin 5/6 (available originally from Boehringer Mannheim, Indianapolis, IN, cat. no. 1273396 and now from Chemicon International, Temekula, CA), anti-keratin 17 (clone E3, available from Dako, Carpinteria, CA, cat. no. M7046), anti-CD3 (available from Dako), and anti-immunoglobulin light chain (A191, A193, available from Dako). These immunohistochemical methods were applied for all the immunohistochemical studies described in the present application unless otherwise stated.

Example 4

cDNA Synthesis and Labeling and Microarray Hybridization

mRNA was isolated from breast tissue, breast tumor samples, and cell lines as described in Example 2. Fluorescently labeled cDNA was synthesized from the mRNA using a reverse transcriptase reaction that included dUTP labeled with either Cy3 or Cy5. For each hybridization experiment differentially labeled cDNA samples (an experimental sample and a reference sample) were pooled and hybridized to a

cDNA microarray, which was then scanned as described in Example 4. The protocol below provides details of the steps performed for cDNA synthesis and labeling and for microarray hybridization.

1. To set up for the reverse transcriptase (RT) reaction, combine the following (e.g., in an Eppendorf tube):

(a) Anchored Oligo dT primer - 2 microliters at 2.5 micrograms/microliter or control - 2 microliters

(b) mRNA - (whatever volume is needed to reach 1.5-2 micrograms)

(c) DEPC/H2O - add sufficient volume so that final volume is 16 microliters

2. Heat at 70° C for 10 minutes
3. Chill on ice for 1-2 minutes
4. Add the following RT reaction components to each individual tube:

(a) 5X RT Buffer - 6 microliters

(b) 50X dNTPs - 0.7 microliters - (500mm A,C,G, 200mm T)

(c) Cy Dyes dUTP - 3 microliters - (either Cy3 or Cy5)

(d) DTT Stock - 3 microliters - (comes with RT setup)

(e) Superscript II RT--1.7 microliters - (cat# 18064-014 Gibco-BRL)

5. Mix well
6. Incubate at 42° C for 1 hour
7. Add another 1 microliter of Superscript II RT and mix
8. Incubate at 42° C for 1 more hour

9. Degrade mRNA with 1.5 microliters of 1M NaOH / 2mM EDTA

10. Incubate at 65° C for 8 minutes (do NOT go TOO long here)

11. Add 15 microliters of 0.1M HCL

12. Add 450 microliters of TE (pH 7.4) to each sample and place each sample into a microcon-30 filter.

13. Add 15 microliters of Human COT1 DNA (Gibco-BRL = 1 microgram/microliter) to each sample in the microcon filter.

14. Spin in Eppendorf centrifuge until volume equals about 50 microliters (8-10')

15. Remove flowthroughs, and pool Cy3 and Cy5 flowthroughs together for future recovery of Cy dyes (store at –20 ° C).

16. Invert microcons, recover labeled samples, and pool Cy3 and Cy5 samples together that will be used for an individual experiment, in a single microcon filter that was used in step 15.

17. Add 500 microliters of T.E again, and spin until final volume equals 8 microliters or less (BE VERY CAREFUL TO NOT SPIN THE SAMPLE DRY!!!)

18. To the 8 microliter combined Cy3 + Cy5 sample, add the following:

(a) Yeast tRNA - 1 microliter - (10 micrograms/microliter)

(b) PolyA DNA - 2 microliters - (10 micrograms/microliter)

(c) 20XSSC - 2 microliters - (FINAL SSC concentration approximately 3X)

(d) 10% SDS - 0.3 microliters

FINAL VOLUME = 13.3 MICROLITERS

19. Mix well.

20. Heat sample at 100° C for 2 minutes, spin very briefly.

21. Place samples at 42° C for 20-30 minutes.

22. During Step 21, prepare the necessary number of hybridization chambers (Custom made by Die-Tech, San Jose, CA (see "Drawings for custom parts at

http://cmgm.stanford.edu/pbrown/mguide/HybChamber.pdf") or purchased at Corning Costar, Acton, MA (CTM™ Hybridization Chamber, #2551), get 22mm X 22mm coverslips ready, and get arrays ready.

23. Add the 13 microliters of probe (i.e., labeled cDNA mixture) onto the center of the array while NOT actually touching the array face with the pipette tip.

24. Quickly and gently place the 22mm X 22mm glass#1 coverslip onto the array face.

25. Add about 15-20 microliters of 3XSSC in two drops onto the end of the array slide away from the actual array for hydration purposes.

26. Assemble the hybridization chamber with the array slide in it, and place into a 65 C water bath overnight.

27. Pull out the hybridization chamber and dry off the excess H_2O.

28. Disassemble the hybridization chamber, and quickly place the slides into a slide washing chamber that contains 2XSSC/0.05%SDS. Jiggle the slide holder up and down until the slide coverslip falls off. Repeat this individually for each array, one at a time, until all are done

29. Wash slides in 1XSSC for 3-5 minutes.

30. Wash slides in 50 C 0.2XSSC for 3-5 minutes, twice.

31. Spin slides down in centrifuge at 200 RPM for 2 minutes.

32.SCAN immediately.

Example 5

Collection, Processing, and Analysis of Data from Microarray Hybridizations

The cDNA microarrays were scanned with either a General Scanning (Watertown, MA) ScanArray 3000 at 20 microns resolution, or with a prototype Axon Instruments (Foster City, CA) GenePix Scanner at 10 micron resolution. The output files, which were TIFF images, were then analyzed using the program ScanAlyze (M. Eisen; available at http://www.microarrays.org/software). Fluorescent ratios and quantitative data on spot quality (see ScanAlyze manual) were stored in a prototype of the AMAD database (M. Eisen; available at http://www.microarrays.org/software). Areas of the array with obvious blemishes were manually flagged and excluded from

subsequent analyses. The primary data tables can be downloaded at http://genome-www.stanford.edu/molecularportraits/, in text/tab delimited format after obtaining a password.

Data were extracted from the database in a single table, with each row representing an array element, each column a hybridization, and each cell the observed fluorescent ratio for the array element in the appropriate hybridization. Previously flagged spots were excluded, as were spots that did not pass quality control. This table had 9216 rows and 84 columns. Array elements were removed if they were not well measured in at least 80% of the hybridizations. The data table was split into tumors and cell lines, and the two subtables were separately median polished (the rows and columns were iteratively adjusted to have median 0) before being rejoined into a single table. Genes whose expression varied by at least 4-fold from the median in this sample set in at least three of the samples tested were selected for the analyses that led to the identification of polynucleotides encoding BSTP-ECG1 (1753 genes satisfied these conditions).

Average-linkage hierarchical clustering, as implemented in the program Cluster (M. Eisen; http://www.microarrays.org/software), was applied separately to both the genes and arrays. The results were analyzed, and figures generated, using TreeView (M. Eisen; http://www.microarrays.org/software).

Example 6

Producing Antibodies to the BSTP-ECG1 Polypeptide

This example describes the preparation of a polyclonal antibody that binds to the BSTP-ECG1 polypeptide, i.e., an antibody that binds to a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 1. The example further describes affinity purification of the antibody.

Materials

- Anisole (Cat. No. A4405, Sigma)

- 2,2'-azino-di-(3-ethyl-benzthiazoline-sulfonic acid) (ABTS) (Cat. No. A6499, Molecular Probes Eugene, OR)
- Activated Maleimide Keyhole Limpet Cyanin (Cat. No. 77106, Pierce Chemical Co. Rockford, IL)
- Biotin (Cat. No. B2643, Sigma)
- Boric acid (Cat. No. B0252, Sigma)
- Sepharose 4b (Cat. No. 17-0120-01, LKB/Pharmacia, Uppsala, Sweden)
- Bovine Serum Albumin (LP) (Cat. No. 100 350, Boehringer Mannheim, Indianapolis, IN)
- Cyanogen bromide (Cat. No. C6388 Sigma, St. Louis, MO)
- Dialysis tubing Spectra/Por Membrane MWCO: 6-8,000 (Cat. No. 132 665, Spectrum Industries Inc., Laguna Hills, CA)
- Dimethyl formamide (DMF) (Cat. No. 22705-6, Aldrich Chemical Company, Milwaukee, WI)
- DIC (Cat. No. BP 592-500, Fisher)
- Ethanedithiol (Cat. No. 39,802-0, Aldrich Chemicals, Milwaukee, WI)
- Ether (Cat. No. TX 1275-3, EM Sciences)
- Ethylenediaminetetraacetatic acid (EDTA)(Cat No. BP 120-1, Fisher Scientific, Springfield, NJ)
- 1-ethyl-3-(3'dimethylaminopropyl)-carbodiimide, HCL (EDC) (Cat No. 341-006, Calbiochem, San Diego, CA)
- Freund's Adjuvant, complete (Cat. No. M-0638-50B, Lee Laboratories, Grayson, GA)
- Freund's Adjuvant, incomplete (Cat. No. M0639-50B, Lee Laboratories)
- Fritted chromatography columns (Column part No. 12131011; Frit: Part No. 12131029, Varian Sample Preparation Products, Harbor City, CA)
- Gelatin from Bovine Skin (Cat. No. G9382, Sigma)
- Glycine (Cat. No. BP381-5, Fisher)
- Goat anti-rabbit IgG, biotinylated (Cat No. A 0418, Sigma)
- HOBt (Cat. No. 01-62-0008, Calbiochem-Novabiochem)
- Horseradish peroxidase (HRP) (Cat. No. 814 393, Boehringer Mannheim)
- HRP-Streptavidin (Cat. No. S 5512, Sigma)

- Hydrochloric Acid (Cat No. 71445-500, Fisher)
- Hydrogen Peroxide 30% w/w (Cat. No. H1009, Sigma)
- Methanol (Cat. No. A412-20, Fisher)
- Microtiter plates, 96 well (Cat. No. 2595, Corning-Costar Pleasanton, CA)
- N-□-Fmoc protected amino acids available from Calbiochem-Novabiochem, San Diego, CA. See 1997-1998 catalog pages 1-45.
- N-□-Fmoc protected amino acids attached to Wang Resin available from Calbiochem-Novabiochem. See 1997-1998 catalog pages 161-164.
- NMP (Cat. No. CAS 872-50-4, Burdick and Jackson, Muskegon, MI)
- Peptide (Synthesized by Research Genetics, Inc. Details given below)
- Piperidine (Cat. No. 80640, Fluka, available through Sigma)
- Sodium Bicarbonate (Cat. No. BP328-1, Fisher)
- Sodium Borate (Cat. No. B9876, Sigma)
- Sodium Carbonate (Cat. No. BP357-1, Fisher)
- Sodium Chloride (Cat. No. BP 358-10, Fisher)
- Sodium Hydroxide (Cat. No. SS 255-1, Fisher)
- Streptavidin (Cat. No. 1 520, Boehringer Mannheim)
- Thioanisole (Cat. No. T-2765, Sigma)
- Trifluoroacetic acid (Cat. No. TX 1275-3, EM Sciences)
- Tween-20 (Cat. No. BP 337-500, Fisher)
- Wetbox-(Rubbermaid Rectangular Servin' Saver™ Part No. 3862 Wooster, OH)

Solutions

- BBS - Borate Buffered Saline with EDTA dissolved in distilled water (pH 8.2 to 8.4 with HCl or NaOH)
  -25 mM Sodium borate (Borax)
  -100 mM Boric Acid
  -75 mM NaCl
  -5 mM EDTA
- 0.1 N HCl in saline
  -concentrated HCl (8.3 mL/0.917 L distilled water)
  -0.154 M NaCl

- Glycine (pH 2.0 and pH 3.0) dissolved in distilled water and adjusted to the desired pH.
  - -0.1 M glycine
  - -0.154 M NaCl
- 5X Borate 1X Sodium Chloride dissolved in distilled water.
  - -0.11 M NaCl
  - -60 mM Sodium Borate
  - -250 mM Boric Acid
- Substrate Buffer in distilled water adjusted to pH 4.0 with sodium hydroxide:
  - -50 to 100 mM Citric Acid

Peptide Synthesis Solutions

- AA solution: HOBt is dissolved in NMP (8.8 grams HOBt to 1 liter NMP). Fmoc-N-a-amino at a concentration at .53 M.
- DIC solution: 1 part DIC to 3 parts NMP.
- Deprotecting solution: 1 part Piperidine to 3 parts DMF
- Reagent R: 2 parts anisole, 3 parts ethanedithiol, 5 parts thioanisole, 90 parts trifluoroacetic acid.

Equipment

- MRX Plate Reader (Dynatech Inc., Chantilly, VA)
- Hamilton Eclipse (Hamilton Instruments, Reno, NV)
- Beckman TJ-6 Centrifuge, Refrigerated (Model No. TJ-6, Beckman Instruments, Fullerton, CA)
- Chart Recorder (Recorder 1 Part No. 18-1001-40, Pharmacia LKB Biotechnology)
- UV Monitor (Uvicord SII Part No. 18-1004-50, Pharmacia LKB Biotechnology)
- Amicon Stirred Cell Concentrator (Model 8400, Amicon Inc., Beverly, MA)
- 30 kD MW cut-off filter (Cat. No. YM-30 Membranes Cat. No. 13742, Amicon Inc., Beverly, MA)
- Multi-channel Automated Pipettor (Cat. No. 4880, Corning Costar Inc., Cambridge, MA)
- pH Meter Corning 240 (Corning Science Products, Corning Glassworks, Corning,

NY)

- ACT396 peptide synthesizer (Advanced ChemTech, Louisville, KY)
- Vacuum dryer (Box is from Labconco, Kansas City, MO; Pump is from Alcatel, Laurel MD).
- Lyophilizer (Unitop 600sl in tandem with Freezemobile 12, both from Virtis, Gardiner, NY)

Methods

Peptides were selected using the program Omiga ™1.1 (Oxford Molecular Group, Inc., 2105 So. Bascom Ave., Suite 200, Campbell, CA 95008) using the Hopp/Woods method, which is described in Hopp TP, Woods KR, *Mol Immunol,* Apr;20(4):483-9 A computer program for predicting protein antigenic determinants, 1983, and Hopp TP and Woods KR, *Proc. Nat. Acad. Sci.* U.S.A. 78, 3824-3828, 1981. Three peptide sequences were selected. The sequences were selected from regions of the polypeptide that displayed minimal homology with known proteins. The sequences of the three peptides were as follows:

Peptide 1 (SEQ ID NO: 6): KYIGFAPCIFHGRGLFSS
Peptide 2 (SEQ ID NO: 7): ESLSSMPGKNAVTLR
Peptide 3 (SEQ ID NO: 8): NRKGFVKLALRHGAD

Synthesis of Peptides

Each of the three peptides listed above was synthesized according to the following protocol:

Incubate: Resin was immersed in appropriate solution. All incubation steps occurred with mixing.

Wash: Added 2 mls. DMF, incubated 5 minutes and drained.

Wash Cycle: Five washes.

Machine Synthesis

The sequence of the desired peptide was provided to the peptide synthesizer. The C-terminal residue was determined and the appropriate Wang Resin was attached to the

reaction vessel. The peptides were synthesized C-terminus to N-terminus by adding one amino acid at a time using a synthesis cycle. Which amino acid is added was controlled by the peptide synthesizer, which looks to the sequence of the peptide entered into its database.

Step 1 – Resin Swelling: Added 2 mL DMF, incubated 30 minutes, drained DMF.
Step 2 – Synthesis cycle
2a – Deprotection: 1 mL deprotecting solution was added to the reaction vessel and incubated for 20 minutes.
2b – Wash Cycle
2c – Coupling: 750 mL of amino acid solution and 250 mL of DIC solution were added to the reaction vessel. The reaction vessel was incubated for thirty minutes and washed once. The coupling step was repeated once.
2d – Wash Cycle
Step 2 was repeated over the length of the peptide. The amino acid solution changed as the sequence listed in peptide synthesizer dictated.
Step 3 – Final Deprotection: Steps 2a and 2b were performed one last time.

Resins were deswelled in methanol—rinsed twice in 5 mL methanol, incubated 5 minutes in 5 mL methanol, rinsed in 5 mL methanol—and then vacuum dried.

Peptide was removed from the resin by incubating 2 hours in reagent R and then precipitated into ether. Peptide was washed in ether and then vacuum dried. Peptide was resolubilized in diH20, frozen, and lyophilized overnight.

Conjugation of Peptide with Keyhole Limpet Hemocyanin

Peptide (6 mg) was dissolved in PBS (6 mL) and mixed with 6 mg of maleiimide activated KLH carrier in 6 mL of PBS for a total volume of 12 mL. The entire solution was mixed for two hours, dialyzed in 1L PBS, and lyophilized.

Immunization of Rabbits

Two New Zealand White Rabbits were injected with 250 μg keyhole limpet

hemocyanin (KLH) conjugated peptide in an equal volume of complete Freund's adjuvant and saline in a total volume of 1 mL. Antigens (KLH-Peptide, 100 μg each) in an equal volume of incomplete Freund's Adjuvant and saline were injected into three to four subcutaneous dorsal sites for a total volume of 1 mL two, four, and six weeks after the first immunization. The three peptides were injected together.

The immunization schedule was as follows:

Day 0	Pre-immune bleed, primary immunization
Day 15	1st Boost
Day 27	1st Bleed
Day 44	2nd Boost
Day 57	2nd Bleed and 3rd Boost
Day 69	3rd Bleed
Day 84	4th boost
Day 98	4th bleed

The Collection of Rabbit Serum

The rabbits were bled (30 to 50 mL) from the auricular artery. The blood was allowed to clot at room temperature for 15 minutes and the serum was separated from the clot using an IEC DPR-6000 centrifuge at 5000 x g. Cell-free serum was decanted gently into a clean test tube and stored at -20°C for affinity purification.

Determination of Antibody Titer

All solutions with the exception of wash solution were added by the Hamilton Eclipse, a liquid handling dispenser. The antibody titer was determined in the rabbits using an ELISA assay with peptide on the solid phase. Flexible high binding ELISA plates were passively coated with peptide diluted in BBS (100 μL, 1 μg/well) and the plate was incubated at 4°C in a wetbox overnight (air-tight container with moistened cotton balls). The plates were emptied and then washed three times with BBS containing 0.1% Tween-20 (BBS-TW) by repeated filling and emptying using a semi-automated plate washer. The plates were blocked by completely filling each well with

BBS-TW containing 1% BSA and 0.1% gelatin (BBS-TW-BG) and incubating for 2 hours at room temperature. The plates were emptied and sera of both pre- and post-immune serum were added to wells. The first well contained sera at 1:50 in BBS. The sera were then serially titrated eleven more times across the plate at a ratio of 1:1 for a final (twelfth) dilution of 1:204,800. The plates were incubated overnight at 4°C. The plates were emptied and washed three times as described.

Biotinylated goat anti-rabbit IgG (100 μL) was added to each microtiter plate test well and incubated for four hours at room temperature. The plates were emptied and washed three times. Horseradish peroxidase-conjugated Streptavidin (100 μL diluted 1:10,000 in BBS-TW-BG) was added to each well and incubated for two hours at room temperature. The plates were emptied and washed three times. The ABTS was prepared fresh from stock by combining 10 mL of citrate buffer (0.1 M at pH 4.0), 0.2 mL of the stock solution (15 mg/mL in water) and 10 μL of 30% H_2O_2. The ABTS solution (100μL) was added to each well and incubated at room temperature. The plates were read at 414 λ, 20 minutes following the addition of substrate.

Preparation of the Peptide Affinity Purification Column:

The affinity column was prepared by conjugating 5 mg of peptide to 10 mL of cyanogen bromide-activated Sepharose 4B, and 5 mg of peptide to hydrazine-Sepharose 4B. Briefly, 100 uL of DMF was added to peptide (5 mg) and the mixture was vortexed until the contents were completely wetted. Water was then added (900 μL) and the contents were vortexed until the peptide dissolved. Half of the dissolved peptide (500 μL) was added to separate tubes containing 10 mL of cyanogen-bromide activated sepharose 4B in 0.1 mL of borate buffered saline at pH 8.4 (BBS), and 10 mL of hydrazine-Sepharose 4B in 0.1 M carbonate buffer adjusted to pH 4.5 using excess EDC in citrate buffer pH 6.0. The conjugation reactions were allowed to proceed overnight at room temperature. The conjugated sepharose was pooled and loaded onto fritted columns, washed with 10 mL of BBS, blocked with 10 mL of 1 M glycine, and washed with 10 mL 0.1 M glycine adjusted to pH 2.5 with HCl and re-neutralized in BBS. The column was washed with enough volume for the optical density at 280λ to reach baseline.

Affinity Purification of the Antibody

The peptide affinity column was attached to a UV monitor and chart recorder.

The titered rabbit antiserum was thawed and pooled. The serum was diluted with one volume of BBS and allowed to flow through the columns at 10 mL per minute. The non-peptide immunoglobulins and other proteins were washed from the column with excess BBS until the optical density at 280 λ reached baseline. The columns were disconnected and the affinity purified column was eluted using a stepwise pH gradient from pH 7.0 to pH 1.0. The elution was monitored at 280 nM, and fractions containing antibody (pH 3.0 to pH 1.0) were collected directly into excess 0.5 M BBS. Excess buffer (0.5 M BBS) in the collection tubes served to neutralize the antibodies collected in the acidic fractions of the pH gradient.

The entire procedure was repeated with "depleted" serum to ensure maximal recovery of antibodies. The eluted material was concentrated using a stirred cell apparatus and a membrane with a molecular weight cutoff of 30 kD. The concentration of the final preparation was determined using an optical density reading at 280 nM. The concentration was determined using the following formula: $mg/mL = OD_{280}/1.4$.

Example 7

SDS-PAGE and Immunoblot Analysis of BSTP-ECG1

To investigate the expression pattern of BSTP-ECG1, extracts are made from a variety of different cell lines and subjected to SDS-PAGE followed by immunoblotting according to the protocol below, using an affinity purified polyclonal antibody to BSTP-ECG1 prepared as described in Example 6.

Materials

- Acetic acid, Glacial (Cat. No. A38[c]-212, Fisher)
- Acrylamide (Cat. No. A-3553, Sigma)
- Anti-Rabbit IgG (H&L) (Cat. No. 31460ZZ, Pierce)
- Bis-acrylamide (Cat. No. M-7279, Sigma)

- Blotting paper (Cat. No. 170-3960, Bio-Rad, Hercules, CA)
- Bovine Serum Albumin (LP) (Cat. No. 100-350, Boehringer Mannheim, Indianapolis, IN)
- Brilliant Blue R-250 (Cat. No. BP101-25, Fisher)
- Complete™ Mini (Cat. No. 1836153, Boehringer Mannheim)
- ECL Western Blotting Detection Reagents (Cat. No. RPN2106, Amersham Pharmacia Biotech, Piscataway, NJ)
- Ethyl alcohol (AAPER Alcohol and Paper Chemical Co., Shelbyville, KY)
- Gelplate Clean (Cat. No. 786-140RF, Geno Technology, Inc., St. Louis)
- Gelatin (Cat. No. G-2500, Sigma)
- Glycerol (Cat. No. BP229-1, Fisher)
- Glycine (Cat. No. G-8898, Sigma)
- Hybond ECL (Cat. No. RPN303D, Amersham Pharmacia Biotech)
- Lauryl Sulfate (SDS) (Cat. No. L-3771, Sigma)
- Methanol (Cat. No. BP1105-4, Fisher)
- M-Per (Cat. No. 78501, Pierce, Rockford, IL)
- Nalgene bottle top filters (Cat. No. 09-740-62B, Fisher)
- Nonfat dry milk (Kroger Co., Cincinnati, OH)
- Ponceau-S (Cat. No. P-07170, Sigma)
- Potassium phosphate (Cat. No. P-0662, Sigma)
- 2X SDS gel loading buffer (Cat. No. 750006, Research Genetics, Huntsville, AL)
- Size markers (Cat. No. M-3913, M-4038, M-3788, Sigma)
- Sodium azide (Cat. No. S227I-25, Fish)
- Sodium chloride (Cat. No. S271-3, Fisher)
- Sodium phosphate, Dibasic, Anhydrous (Cat. No. BP332-1, Fisher)
- t-amyl alcohol (Cat. No. A-16852, Sigma)
- TEMED (Cat. No. T-9281, Sigma)
- Trizma® Base (Cat. No. T-6066, Sigma)
- Tween-20 (Cat. No. BP337-500, Fisher)

Solutions

- PBS – Phosphate Buffered Saline dissolved in distilled water

-136 mM NaCl

-2.7 mM KCl

-10.1 mM Na_2HPO_4

-1.8 mM KH_2PO_4

- Acrylamide/Bis (30% T, 2.67% C) dissolved in distilled water

  -4.1 M acrylamide

  -51.9 mM N,N'-
- 1.5 M Tris-HCl (pH 8.8) dissolved in distilled water
- 0.5 M Tris-HCl (pH 6.8) dissolved in distilled water
- 10% SDS – dissolve 10 grams SDS in 100 mls distilled water
- Running Buffer

  -24.8 mM Tris base

  -191.9 mM glycine

  -3.5 mM SDS
- Towbin transfer buffer (pH 8.3) dissolved in distilled water

  -20% methanol

  -25 mM Tris

  -192 mM glycine
- Equilibrating buffer for gel drying, mixed in distilled water

  -20% ethanol

  -10% glycerol
- Gel staining solution dissolved in distilled water

  -0.3 mM Coomassie brilliant blue R-250

  -40% methanol

  -7% glacial acetic acid
- Gel destaining solution mixed in distilled water

  -25% methanol

  -7% glacial acetic acid
- 10% Tween®20 in PBS
- 5% Nonfat dry milk in PBS
- 0.2% BSA Blocking Buffer dissolved in PBS

  -0.2% BSA

-0.1% gelatin

-0.05% Tween®20

- Wash Buffer

-0.05% Tween®20

-1X PBS

Equipment

- Microcentrifuge (Model 5415, Eppendorf)
- Power Pak 200 (Cat. No. 165-5052, Bio-Rad)
- Power Pak 3000 (Cat. No. 165-5056, Bio-Rad)
- Protean II xi Cell (Cat. No. 165-1813, Bio-Rad)
- Recirculating chiller (Cat. No. CFT33D115V, Neslab Instruments, Inc., Portsmouth, NH)
- 20-Well comb (Cat. No. 165-1867, Bio-Rad)
- pH Meter Corning 240 (Corning Science Products, Corning Glasswares, Corning, NY)
- Air Cadet vacuum pump (Cat. No. P-07530-50, Cole-Palmer Instruments Co., Chicago, IL)
- Tissue Tearor tissue homogenizer (Cat. No. 985370-07, BioSpec Products Inc., Bartletsville, OK)

Methods

Sample Preparation

Preferably a variety of cell lines known in the art are used for the experiment, including cell lines derived from breast tumors, cell lines derived from normal breast tissue, cell lines derived from other cancer types, etc. A selection of appropriate cancer cell lines for investigation of the expression of BSTP-ECG1 is found in reference 21. A selection of noncancer cell lines appropriate for investigation of the expression of BSTP-ECG1 is found in Perou, et al., Molecular portraits of human breast tumours, *Nature*, 406(6797):747-52, 2000. Appropriate cell lines include MCF7, Hs578T, OVCAR3, HepG2, NTERA2, MOLT4, RPMI-8226, NB4+ATRA, UACC-62, SW872, and Colo205: also see Table 2 for more details). Cell lines are maintained under standard growth conditions and in standard tissue culture media as

appropriate for the particular cell line. Cells are collected according to standard techniques (e.g., trypsinization in the case of adherent cells), and the resulting cell suspension is prepared as follows:

-The cell suspension is pelleted by centrifugation at 3000 RPM for 10 minutes, and the supernatant was discarded.

-The pellet is washed with 1ml PBS, centrifuged at 10000 RPM for 10 minutes, and the supernatant was discarded.

-An appropriate volume of M-Per™ Reagent is added to the cell pellet and mixed gently for 10 minutes in an ice bath. The mixture is centrifuged at 13200 RPM for 15 minutes, and the supernatant is saved.

The protein concentration in the supernatant is measured according to standard techniques.

All samples are mixed at 1:1 with gel loading buffer and boiled for 5 minutes before loading.

SDS PAGE

Standard SDS-PAGE stacking and running gels are prepared and placed in an electrophoresis apparatus. After filling the upper and lower chambers with running buffers the samples (60 □g/lane) are loaded. The inner core is placed in the lower chamber and the lid placed on top. The apparatus is connected to the power supply and recirculating system. The temperature setting is 10°C. The stacking gel is run at 14mA per gel for 1 hour. The separating gel is run at 0.58mA per gel per hour for 16 hours.

Transfer to nitrocellulose

After electrophoresis is complete, the gel is equilibrated in Towbin Buffer for 15-30 minutes. The assembly for transfer is as follows:

cathode
pre-soaked blotting paper
gel
pre-wetted nitrocellulose
pre-soaked blotting paper

anode

The transfer is performed at 20V for 25 minutes, then 25V for 20 minutes. After the transfer is complete, the gel is stained with Coomassie and the blot is stained with Ponceau-S.

Western Blotting

Primary and secondary antibodies

All primary and secondary antibodies are diluted in 0.2% BSA blocking buffer. All incubation steps are done with gentle mixing.

Blots are blocked in 5% milk overnight at room temperature. The blots are rinsed with wash buffer before adding the primary antibody and incubating for two hours at room temperature.

One wash cycle is performed. One wash cycle consists of:

Wash 5 min, rinse
Wash 5 min, rinse
Wash 10 min, rinse
Wash 5 min, rinse
Wash 5 min, rinse

The secondary antibody is added and incubated for one hour at room temperature. One wash cycle is then performed.

Peptide Block

As a control to demonstrate the specificity of the antibody, equal amounts (w/w) of peptide and antibody are added to 1/10 of the final volume of blocking buffer and incubated overnight at 4°C. The volume of blocking buffer is then brought up to the final volume, and the membrane is incubated for an additional two hours at room temperature.

Developing

The blots are placed in a Ziploc® bag. Equal volumes of ECL western blotting detection reagents are mixed and distributed evenly over the blots. The blots are placed in an autoradiography cassette, covered with a piece of film, and exposed.

Example 8

Detection of BSTP-ECG1 in Breast Cancer Samples by Immunohistochemistry

A breast cancer tissue microarray consisting of tissue samples from a large number of breast cancer biopsies is prepared essentially as described in Kononen, J., et al., Tissue microarrays for high-throughput molecular profiling of tumor specimens, *Nature Medicine*, 4(7), 844-847, 1998. Briefly, several hundred archival paraffin-embedded breast tumor samples were obtained from the Pathology Department at Stanford University Medical Center. The samples were reviewed by a pathologist (applicant MVR) to ensure that they met pathological criteria for breast cancer. Small tissue cores were removed from the samples and embedded in a single paraffin block to produce a tissue array. Immunohistochemistry is performed as described previously (Perou, C., *et al.*, 1999; Bindl, J. and Warnke, R., *Am J Clin Pathol*, 85, 490-493, 1986, and Natkunam, Y., *et al.*, Am. J. Path., 156(1), 2000, the contents of which are incorporated herein by reference) using an antibody to BSTP-ECG1 generated as described in Example 6.

Example 10

Expression of BST-ECG1 mRNA in Cell Lines

Materials and Methods

Cell lines were grown from frozen stocks in RPMI-1640 supplemented with 10% fetal bovine serum. Total RNA was extracted at approximately 80-90% confluence using TRIZOL reagent (Life Technologies) according to the manufacturer's protocol. Following total RNA extraction, mRNA was isolated with the FastTrack 2.0 kit (Invitrogen) following the manufacturer's protocol for isolating mRNA from total RNA. HepG2 is a liver tumor derived cell line (ATCC #HB-8065); COLO205 is a colon tumor derived cell line (ATCC #CCL-222); and MCF-7 is a breast adenocarcinoma derived cell line (ATCC #HTB-22).

Five micrograms of each mRNA and RNA ladder ranging from 0.24-9.49kb (Life

Technologies) was size separated and blotted onto Hybond N+ membrane (Amersham). The membrane was hybridized at 42°C to a 32P-labeled cDNA probe of the *BST-ECG1* gene in Hybrisol (Oncor) with 20□g Sheared DNA (Research Genetics) and 5□g human Cot1 DNA (Life Technologies). Following hybridization, the nylon membrane was exposed to a phosphor screen. The digital image of the Northern was then acquired using the Packard cyclone phosphor imager in order to assess the size of the *BST-ECG1* transcript.

Results

The Northern blot showed 2 bands of approximately 1.5 and 2.2 kB, consistent with the prediction of multiple isoforms due to alternate 3' processing (alternate polyadenylation sites). Expression of *BST-ECG1* is present in all three cell types tested (Figure 6A). The highest level of expression was observed in the HepG2 cell line. Figure 6B presents a longer exposure demonstrating expression in MCF-7 cells.

REFERENCES

1. Tavassoli, F. A. & Schnitt, S. J. Pathology of the breast (Elsevier, New York, 1992).

2. Fambrough, D., McClure, K., Kazlauskas, A. & Lander, E. S. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes [see comments]. Cell 97, 727-741 (1999).

3. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S. & Fink, G. R. Ploidy regulation of gene expression [see comments]. Science 285, 251-254 (1999).

4. Cho, R. J. et al. A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2, 65-73 (1998).

5. Iyer, V. R. et al. The transcriptional program in the response of human fibroblasts to serum [see comments]. Science 283, 83-87 (1999).

6. Chu, S. et al. The transcriptional program of sporulation in budding yeast [published erratum appears in Science 1998 Nov 20;282(5393):1421]. Science 282, 699-705 (1998).

7. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680-686 (1997).

8. Perou, C. M. et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 96, 9212-9217 (1999).

9. Alon, U. et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A 96, 6745-6750 (1999).

10. Alizadeh, A. A. et al. Distinct Types of Diffuse Large B-Cell Lymphoma Identified by Gene Expression Profiling. Nature 403(6769):503-11, 2000.

11. Golub, T. R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531-537 (1999).

12. Khan, J. et al. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res 58, 5009-5013 (1998).

13. Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Breast Cancer Linkage Consortium [see comments]. Lancet 349, 1505-1510 (1997).

14. Andersen, T. I. et al. Prognostic significance of TP53 alterations in breast

carcinoma. Br J Cancer 68, 540-548 (1993).

15. Aas, T. et al. Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. Nat Med 2, 811-814 (1996).

16. Slamon, D. J. et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244, 707-712 (1989).

17. Osborne, C. K., Yochmowitz, M. G., Knight, W. A. d. & McGuire, W. L. The value of estrogen and progesterone receptors in the treatment of breast cancer. Cancer 46, 2884-2888 (1980).

18. Ronnov-Jessen, L., Petersen, O. W. & Bissell, M. J. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. Physiol Rev 76, 69-125 (1996).

19. Eisen, M. B. & Brown, P. O. DNA arrays for analysis of gene expression. Methods Enzymol 303, 179-205 (1999).

20. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95, 14863-14868 (1998).

21. Ross, D. T. et al. Systematic Variation in Gene Expression Patterns in Human Cancer Cell Lines. Nature Genetics, 24(3):227-35, 2000.

22. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. Nature 396, 643-649 (1998).

23. Cahill, D. P. et al. Mutations of mitotic checkpoint genes in human cancers [see comments]. Nature 392, 300-303 (1998).

24. Li, Y. & Benezra, R. Identification of a human mitotic checkpoint gene: hsMAD2. Science 274, 246-248 (1996).

25. Zhou, H. et al. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation [see comments]. Nat Genet 20, 189-193 (1998).

26. Wolf, G. et al. Prognostic significance of polo-like kinase (PLK) expression in non- small cell lung cancer. Oncogene 14, 543-549 (1997).

27. Yang, G. P., Ross, D. T., Kuang, W. W., Brown, P. O. & Weigel, R. J. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes [In Process Citation]. Nucleic Acids Res 27, 1517-1523 (1999).

28. Hoch, R. V., Thompson, D. A., Baker, R. J. & Weigel, R. J. GATA-3 is expressed in association with estrogen receptor in breast cancer. Int J Cancer 84, 122-128 (1999).

29. Prud'homme, J. F. et al. Cloning of a gene expressed in human breast cancer and regulated by estrogen in MCF-7 cells. Dna 4, 11-21 (1985).

30. Bhargava, V., Kell, D. L., van de Rijn, M. & Warnke, R. A. Bcl-2 immunoreactivity in breast carcinoma correlates with hormone receptor positivity. Am J Pathol 145, 535-540 (1994).

31. Leek, R. D., Kaklamanis, L., Pezzella, F., Gatter, K. C. & Harris, A. L. bcl-2 in normal human breast and carcinoma, association with oestrogen receptor-positive, epidermal growth factor receptor-negative tumours and in situ cancer. Br J Cancer 69, 135-139 (1994).

32. Pauletti, G., Godolphin, W., Press, M. F. & Slamon, D. J. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. Oncogene 13, 63-72 (1996).

33. Pollack, J. R. et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 23, 41-46 (1999).

34. Stein, D. et al. The SH2 domain protein GRB-7 is co-amplified, overexpressed and in a tight complex with HER2 in breast cancer. Embo J 13, 1331-1340 (1994).

35. Moog-Lutz, C. et al. MLN64 exhibits homology with the steroidogenic acute regulatory protein (STAR) and is over-expressed in human breast carcinomas. Int J Cancer 71, 183-191 (1997).

36. Miettinen, M., Lindenmayer, A. E. & Chaubal, A. Endothelial cell markers CD31, CD34, and BNH9 antibody to H- and Y- antigens--evaluation of their specificity and sensitivity in the diagnosis of vascular tumors and comparison with von Willebrand factor. Mod Pathol 7, 82-90 (1994).

37. Abbas, A. K., Lichtman, A. H. & Pober, J. S. Cellular and molecular immunology (Saunders, Philadelphia, 1991).

38. Baxa, C. A. et al. Human adipocyte lipid-binding protein: purification of the protein and cloning of its complementary DNA. Biochemistry 28, 8683-8690 (1989).

39. Tontonoz, P., Hu, E., Graves, R. A., Budavari, A. I. & Spiegelman, B. M. mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. Genes Dev 8,

1224-1234 (1994).

40. Stampfer, M.

41. Dairkee, S. H., Mayall, B. H., Smith, H. S. & Hackett, A. J. Monoclonal marker that predicts early recurrence of breast cancer [letter]. Lancet 1, 514 (1987).

42. Dairkee, S. H., Puett, L. & Hackett, A. J. Expression of basal and luminal epithelium-specific keratins in normal, benign, and malignant breast tissue. J Natl Cancer Inst 80, 691-695 (1988).

43. Malzahn, K., Mitze, M., Thoenes, M. & Moll, R. Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas. Virchows Arch 433, 119-129 (1998).

44. Guelstein, V. I. et al. Monoclonal antibody mapping of keratins 8 and 17 and of vimentin in normal human mammary gland, benign tumors, dysplasias and breast cancer. Int J Cancer 42, 147-153 (1988).

45. Gusterson, B. A. et al. Distribution of myoepithelial cells and basement membrane proteins in the normal breast and in benign and malignant breast diseases. Cancer Res 42, 4763-4770 (1982).

46. Nagle, R. B. et al. Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. J Histochem Cytochem 34, 869-881 (1986).

47. Dairkee, S. H., Ljung, B. M., Smith, H. & Hackett, A. Immunolocalization of a human basal epithelium specific keratin in benign and malignant breast disease. Breast Cancer Res Treat 10, 11-20 (1987).

48. Berns, E. M. et al. Prevalence of amplification of the oncogenes c-myc, HER2/neu, and int-2 in one thousand human breast tumours: correlation with steroid receptors. Eur J Cancer 28, 697-700 (1992).

49. Heintz, N. H., Leslie, K. O., Rogers, L. A. & Howard, P. L. Amplification of the c-erb B-2 oncogene and prognosis of breast adenocarcinoma. Arch Pathol Lab Med 114, 160-163 (1990).

50. Guerin, M., Barrois, M., Terrier, M. J., Spielmann, M. & Riou, G. Overexpression of either c-myc or c-erbB-2/neu proto-oncogenes in human breast carcinomas: correlation with poor prognosis. Oncogene Res 3, 21-31 (1988).

51. Wang, K. et al. Monitoring gene expression profile changes in ovarian

carcinomas using cDNA microarray [In Process Citation]. Gene 229, 101-108 (1999).

52. Spellman, PT, et al., Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization, *Mol Biol Cell*, 9(12):3273-97, 1998.

CLAIMS

We claim:

1. A substantially purified polypeptide whose sequence comprises the polypeptide sequence of SEQ ID NO: 1.

1a. A fragment of the polypeptide of claim 1, wherein the fragment is at least 50 amino acids in length.

1b. A fragment of the polypeptide of claim 1, wherein the fragment is at least 100 amino acids in length.

1c. A fragment of the polypeptide of claim 1, wherein the fragment is at least 150 amino acids in length.

1d. A variant of the polypeptide of claim 1, wherein the variant includes between 1 and 10 amino acid substitutions, inclusive.

1e. A variant of the polypeptide of claim 1, wherein the variant includes between 11 and 25 amino acid substitutions, inclusive.

1f. A variant of the polypeptide of claim 1, wherein the variant includes between 26 and 50 amino acid substitutions, inclusive.

1g. A variant of the polypeptide of claim 1, wherein the variant includes an addition or substitution of between 1 and 10 amino acids, inclusive.

1h. A variant of the polypeptide of claim 1, wherein the variant includes an addition or substitution of between 11 and 25 amino acids, inclusive.

1i. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 25% of the length of SEQ ID NO:1, or both.

1j. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 50% of the length of SEQ ID NO:1, or both.

1jj. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 60 or a % positive greater than 70 encompassing at least 75% of the length of SEQ ID NO:1, or both.

1k. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 70 or a % positive greater than 80 encompassing at least 25% of the length of SEQ ID NO:1, or both.

1l. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 70 or a % positive greater than 80 encompassing at least 50% of the length of SEQ ID NO:1, or both.

1ll. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 70 or a % positive greater than 80 encompassing at least 75% of the length of SEQ ID NO:1, or both.

1m. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 80 or a % positive greater than 90 encompassing at least 25% of the length of SEQ ID NO:1, or both.

1n. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 80 or a % positive greater than 90 encompassing at least 50% of the length of SEQ ID NO:1, or both.

1nn. A substantially purified polypeptide having significant similarity to a polypeptide whose sequence is set forth in SEQ ID NO: 1, wherein a polypeptide is considered significantly similar if, when the amino acid sequence of the polypeptide is compared with the amino acid sequence of the polypeptide of SEQ ID NO:1 using the BLAST algorithm and the BLOSUM substitution matrix with default parameters, the result is a % identity greater than 80 or a % positive greater than 90 encompassing at least 75% of the length of SEQ ID NO:1, or both.

2. A purified and isolated polynucleotide comprising the polynucleotide sequence of SEQ ID NO: 2.

3. A purified and isolated polynucleotide comprising the polynucleotide sequence of SEQ ID NO: 3.

5. A purified and isolated polynucleotide having a sequence that is complementary to the polynucleotide sequence of claim 2 or 3.

6. An isolated and purified polynucleotide encoding a polypeptide whose amino acid sequence comprises the amino acid sequence of SEQ ID NO:1.

6a. An isolated and purified polynucleotide encoding a polypeptide having significant similarity to the polypeptide having a sequence set forth in SEQ ID NO:1, wherein a polypeptide having significant similarity the polypeptide having a sequence set forth in SEQ ID NO:1 is defined as in claim 1i, 1j, 1k, 1l, 1m, or 1n.

7. An isolated and purified polynucleotide that hybridizes to the polynucleotide of SEQ ID NO:2 under stringent conditions.

7a. An isolated and purified polynucleotide that hybridizes to the polynucleotide of SEQ ID NO:3 under stringent conditions.

7e. An isolated and purified polynucleotide that hybridizes to the polynucleotide of SEQ ID NO:2 under moderately stringent conditions.

7f. An isolated and purified polynucleotide that hybridizes to the polynucleotide of SEQ ID NO:3 under moderately stringent conditions.

7j. An isolated and purified polynucleotide that hybridizes to the polynucleotide of claim 6 under stringent conditions.

7k. An isolated and purified polynucleotide that hybridizes to the polynucleotide of claim 6a under stringent conditions.

7l. An isolated and purified polynucleotide that hybridizes to the polynucleotide of claim 6 under moderately stringent conditions.

7m. An isolated and purified polynucleotide that hybridizes to the polynucleotide of claim 6a under moderately stringent conditions.

10. An expression vector comprising the polynucleotide of claim 6.

10a. An expression vector comprising the polynucleotide of claim 6a.

11. An expression vector comprising a polynucleotide that encodes the fragment of any of claims 1a, 1b, or 1c.

12. An expression vector comprising a polynucleotide that encodes the variant of any of claims 1d, 1e, 1f, 1g, or 1h.

20. A host cell comprising the expression vector of claim 10.

20a. A host cell comprising the expression vector of claim 10a.

30. A method of producing a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 1 and a polypeptide whose sequence has significant similarity to the amino acid sequence of SEQ ID NO:1, the method comprising the steps of:
culturing the host cell of claim 20 under conditions wherein the polypeptide is expressed; and
recovering the polypeptide from the host cell culture.

50. A pharmaceutical composition comprising:
the polypeptide of claim 1, or a polypeptide whose sequence has significant similarity to the amino acid sequence of SEQ ID NO:1; and
a pharmaceutically acceptable carrier.

55. A pharmaceutical composition comprising:
the polynucleotide of claim 6; and
a pharmaceutically acceptable carrier.

60. A purified antibody that specifically binds to the polypeptide of claim 1.

65. A purified antibody that specifically binds to the polypeptide of SEQ ID NO: 1.

65a. A purified antibody that specifically binds to a polypeptide having significant similarity to the polypeptide having a sequence set forth in SEQ ID NO:1, wherein a polypeptide having significant similarity the polypeptide having a sequence set forth in SEQ ID NO:1 is defined as in claim 1i, 1j, 1k, 1l, 1m, or 1n.

66. The antibody of claim 60 or claim 65, wherein the antibody is a polyclonal antibody.

67. The antibody of claim 60 or claim 65, wherein the antibody is a monoclonal antibody.

70. A pharmaceutical composition comprising:
the antibody of claim 60 or claim 65; and
a pharmaceutically acceptable carrier.

80. A method for detecting a polynucleotide that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1 in a biological sample, the method comprising steps of:
(a) hybridizing a nucleic acid complementary to the polynucleotide that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1, or that encodes a fragment or variant of the polypeptide, to at least one nucleic acid in the biological sample, thereby forming a hybridization complex; and
(b) detecting the hybridization complex, wherein the presence of the hybridization complex indicates the presence of a polynucleotide encoding the polypeptide in the biological sample.

81. The method of claim 80, wherein the biological sample comprises breast cancer tissue or cells.

82. The method of claim 80, wherein the biological sample comprises normal breast tissue or cells.

85. A method for detecting a polynucleotide that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1 comprising steps of:
(a) hybridizing a nucleic acid that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:1 to at least one nucleic acid complementary to a nucleic acid in the biological sample, thereby forming a hybridization complex; and
(b) detecting the hybridization complex, wherein the presence of the hybridization complex indicates the presence of the polynucleotide in the biological sample.

86. The method of claim 85, wherein the biological sample comprises breast cancer tissue or cells.

87. The method of claim 85, wherein the biological sample comprises normal breast tissue or cells.

90. A method for detecting a polypeptide whose sequence comprises the amino acid sequence set forth in SEQ ID NO:1 in a biological sample comprising steps of:

(a) contacting the biological sample with the antibody of claim 60 or claim 65; and

(b) determining whether the antibody specifically binds to the sample, the binding being an indication that the sample contains the polypeptide.

91. The method of claim 90, wherein the biological sample comprises breast cancer tissue or cells.

92. The method of claim 90, wherein the biological sample comprises a cell, tissue, blood, urine, serum, ascites, saliva, or another body fluid, secretion, or excretion.

93. The method of claim 90, wherein the determining step comprises performing an enzyme-linked immunosorbent assay.

94. The method of claim 90, wherein the determining step comprises performing immunohistochemistry.

95. The method of claim 90, wherein the determining step comprises contacting an antibody array with the sample.

90a. A method for detecting a polypeptide having significant similarity to the polypeptide whose sequence is set forth in SEQ ID NO:1 in a biological sample comprising steps of:

(a) contacting the biological sample with the antibody of claim 60 or claim 65; and

(b) determining whether the antibody specifically binds to the sample, the binding being an indication that the sample contains the polypeptide.

91a. The method of claim 90a, wherein the biological sample comprises breast cancer tissue or cells.

92a. The method of claim 90a, wherein the biological sample comprises a cell, tissue, blood, urine, serum, ascites, saliva, or another body fluid, secretion, or excretion.

93a. The method of claim 90a, wherein the determining step comprises performing an enzyme-linked immunosorbent assay.

94a. The method of claim 90a, wherein the determining step comprises performing immunohistochemistry.

95a. The method of claim 90a, wherein the determining step comprises contacting an antibody array with the sample.

100. A method for treating or preventing a disorder of cell proliferation, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of any of claims 50, 55, or 70.

150. A method for classifying a disease comprising the steps of:

(a) providing a sample from a subject;

(b) detecting the presence of a polynucleotide that encodes a polypeptide having the sequence of SEQ ID NO:1 within the sample; and

(c) assigning the disease to one of a set of predetermined categories based on detection of the polynucleotide.

151. The method of claim 150, wherein the disease is cancer.

152. The method of claim 150, wherein the disease is breast cancer.

153. The method of claim 150, wherein the detecting step comprises a nucleic acid amplification step.

154. The method of claim 150, wherein the sample comprises a cell, tissue, blood, urine, saliva, ascites, other body fluid, secretion, or excretion.

150a. A method for classifying a disease comprising the steps of:

(a) providing a sample from a subject;

(b) detecting the presence of a polynucleotide that encodes a polypeptide having significant similarity to a polypeptide whose sequence comprises the sequence of SEQ ID NO:1 within the sample; and

(c) assigning the disease to one of a set of predetermined categories based on detection of the polynucleotide.

151a. The method of claim 150a, wherein the disease is cancer.

152a. The method of claim 150a, wherein the disease is breast cancer.

153a. The method of claim 150a, wherein the detecting step comprises a nucleic acid amplification step.

154a. The method of claim 150a, wherein the sample comprises a cell, tissue, blood, urine, saliva, ascites, other body fluid, secretion, or excretion.

155. The method of claim 150, further comprising the step of:

providing diagnostic, prognostic, or predictive information based on the predetermined category assigned in the assigning step.

155a. The method of claim 150a, further comprising the step of:

providing diagnostic, prognostic, or predictive information based on the predetermined category assigned in the assigning step.

156. The method of claim 155 or claim 155a, wherein the disease is breast cancer.

160. A method for classifying a disease comprising the steps of:

(a) providing a sample from a subject;

(b) detecting the presence of a polypeptide comprising the sequence of SEQ ID NO:1 within the sample; and

(c) assigning the disease to one of a set of predetermined categories based on detection of the polynucleotide.

161. The method of claim 160, wherein the disease is cancer.

162. The method of claim 160, wherein the disease is breast cancer.

164. The method of claim 160, wherein the sample comprises a cell, tissue, blood, urine, saliva, ascites, other body fluid, secretion, or excretion.

165. The method of claim 160, further comprising the step of providing diagnostic, prognostic, or predictive information based on the category assigned in the assigning step.

166. The method of claim 160, wherein the detecting step is performed using an antibody that specifically binds to the polypeptide of SEQ ID NO:1.

166a. The method of claim 166, wherein the determining step comprises performing an enzyme-linked immunosorbent assay.

166b. The method of claim 166, wherein the determining step comprises performing immunohistochemistry.

166c. The method of claim 166, wherein the determining step comprises contacting an antibody array with the sample.

170. A method for classifying a disease comprising the steps of:

(a) providing a sample from a subject;

(b) detecting the presence of a polypeptide having significant similarity to a polypeptide comprising the sequence of SEQ ID NO:1 within the sample; and

(c) assigning the disease to one of a set of predetermined categories based on detection of the polynucleotide.

171. The method of claim 170, wherein the disease is cancer.

172. The method of claim 170, wherein the disease is breast cancer.

174. The method of claim 170, wherein the sample comprises a cell, tissue, blood, urine, saliva, ascites, other body fluid, secretion, or excretion.

175. The method of claim 170, further comprising the step of providing diagnostic, prognostic, or predictive information based on the category assigned in the assigning step.

176. The method of claim 170, wherein the detecting step is performed using an antibody that specifically binds to the polypeptide.

176a. The method of claim 176, wherein the determining step comprises performing an enzyme-linked immunosorbent assay.

176b. The method of claim 176, wherein the determining step comprises performing immunohistochemistry.

176c. The method of claim 176, wherein the determining step comprises contacting an antibody array with the sample.

180. A method for obtaining prognostic, diagnostic, or therapeutic information comprising steps of:

(i) obtaining a sample containing cells or tissue from a subject; and

(ii) detecting, within the sample, a mutation in a regulatory or coding region of a gene that encodes the BSTP-ECG1 polypeptide of SEQ ID NO:1.

220. A method of inhibiting the growth of a cell comprising enhancing the level or activity of a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a polypeptide having significant similarity to the polypeptide of SEQ ID NO:1 in the cell.

226. The method of claim 220, wherein the cell is a tumor cell.

250. A method of treating or preventing a tumor comprising steps of:

(i) providing an individual in need of treatment or prevention of a tumor;

(ii) administering a compound that enhances the level or activity of a polypeptide whose sequence comprises the amino acid sequence of SEQ ID NO:1.

260. A method of treating or preventing a tumor comprising steps of:

(i) providing an individual in need of treatment or prevention of a tumor;

(ii) administering a compound that reduces or inhibits the level or activity of a polypeptide whose sequence comprises the amino acid sequence of SEQ ID NO:1.

310. A method of inhibiting the growth of a cell comprising reducing the level or activity of a polypeptide comprising the amino acid sequence of SEQ ID NO:1 in the cell.

400. A diagnostic kit comprising:

an antibody that specifically binds to the polypeptide of SEQ ID NO:1;

instructions for use of the antibody;

and a control sample, wherein the antibody specifically binds to a polypeptide in the control sample.

601. A method of classifying a tumor comprising the steps of:

providing a tumor sample;

detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the sample; and

classifying the tumor as belonging to a tumor subclass based on the results of the detecting step.

605. The method of claim 601, wherein the detecting step comprises detecting the polypeptide.

606. The method of claim 605, wherein the polypeptide is detected by performing immunohistochemical analysis on the sample using an antibody that specifically binds to the polypeptide.

606a. The method of claim 605, wherein the polypeptide is detected by performing an ELISA assay using an antibody that specifically binds to the polypeptide.

606b. The method of claim 605, wherein the polypeptide is detected using an antibody array comprising an antibody that specifically binds to the polypeptide.

606c. The method of claim 605, wherein the detecting step comprises:

detecting modification of a substrate by the polypeptide.

607. The method of claim 601, wherein classifying a tumor comprises:

stratifying a subject having the tumor for a clinical trial.

608. The method of claim 607, wherein the tumor is a breast tumor.

609. The method of claim 601, wherein the tumor is a breast tumor and the tumor subclass is a luminal tumor subclass.

601a. The method of claim 601, further comprising:

providing diagnostic, prognostic, or predictive information based on the classifying step.

605a. The method of claim 605, further comprising:

providing diagnostic, prognostic, or predictive information based on the classifying step.

606aa. The method of claim 605a, wherein the polypeptide is detected by performing immunohistochemical analysis on the sample using an antibody that specifically binds to the polypeptide.

606ab. The method of claim 605a, wherein the polypeptide is detected by performing an ELISA assay using an antibody that specifically binds to the polypeptide.

606ac. The method of claim 605a, wherein the polypeptide is detected using an antibody array comprising an antibody that specifically binds to the polypeptide.

606ad. The method of claim 605a, wherein the detecting step comprises:

detecting modification of a substrate by the polypeptide.

609a. The method of claim 601a, wherein the tumor is a breast tumor and the tumor subclass is a luminal tumor subclass.

601g. The method of claim 601, further comprising:

selecting a treatment based on the classifying step.

605g. The method of claim 605, further comprising:

selecting a treatment based on the classifying step.

606ag. The method of claim 605g, wherein the polypeptide is detected by performing immunohistochemical analysis on the sample using an antibody that specifically binds to the polypeptide.

606bg. The method of claim 605g, wherein the polypeptide is detected by performing an ELISA assay using an antibody that specifically binds to the polypeptide.

606cg. The method of claim 605g, wherein the polypeptide is detected using an antibody array comprising an antibody that specifically binds to the polypeptide.

606dg. The method of claim 605g, wherein the detecting step comprises:

detecting modification of a substrate by the polypeptide.

609g. The method of claim 601g, wherein the tumor is a breast tumor and the tumor subclass is a luminal tumor subclass.

601m. A method of testing a subject comprising the steps of:

providing a sample isolated from a subject;

detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the sample; and

providing diagnostic, prognostic, or predictive information based on the detecting step.

605m. The method of claim 601m, wherein the detecting step comprises detecting the polypeptide.

606m. The method of claim 605m, wherein the polypeptide is detected by performing immunohistochemical analysis on the sample using an antibody that specifically binds to the polypeptide.

606ma. The method of claim 605m, wherein the polypeptide is detected by performing an ELISA assay using an antibody that specifically binds to the polypeptide.

606mb. The method of claim 605m, wherein the polypeptide is detected using an antibody array comprising an antibody that specifically binds to the polypeptide.

606mc. The method of claim 605m, wherein the detecting step comprises:

detecting modification of a substrate by the polypeptide.

609m. The method of any of claim 601m, wherein the sample is selected from the group consisting of:

a blood sample, a urine sample, a serum sample, an ascites sample, a saliva sample, a cell, and a portion of tissue.

610m. The method of claim 601m, wherein the sample is a tumor sample.

611m. The method of claim 610m, wherein the tumor sample is a breast tumor sample.

601r. A method of testing a subject comprising the steps of:

providing a sample isolated from a subject;

detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the sample; and

stratifying the subject for a clinical trial based on the detecting step.

605r. The method of claim 601r, wherein the detecting step comprises detecting the polypeptide.

606r. The method of claim 605r, wherein the polypeptide is detected by performing immunohistochemical analysis on the sample using an antibody that specifically binds to the polypeptide.

606ra. The method of claim 605r, wherein the polypeptide is detected by performing an ELISA assay using an antibody that specifically binds to the polypeptide.

606rb. The method of claim 605r, wherein the polypeptide is detected using an antibody array comprising an antibody that specifically binds to the polypeptide.

606rc. The method of claim 605r, wherein the detecting step comprises:

detecting modification of a substrate by the polypeptide.

609r. The method of claim 601r, wherein the sample is selected from the group consisting of:

a blood sample, a urine sample, a serum sample, an ascites sample, a saliva sample, a cell, and a portion of tissue.

610r. The method of claim 601r, wherein the sample is a tumor sample.

611r. The method of claim 610r, wherein the tumor sample is a breast tumor sample.

601q. A method of testing a subject comprising the steps of:

providing a sample isolated from a subject;

detecting expression or activity of a gene encoding the polypeptide of SEQ ID NO:1 in the sample; and

selecting a treatment based on the detecting step.

605q. The method of claim 601q, wherein the detecting step comprises detecting the polypeptide.

606q. The method of claim 605q, wherein the polypeptide is detected by performing immunohistochemical analysis on the sample using an antibody that specifically binds to the polypeptide.

606qa. The method of claim 605q, wherein the polypeptide is detected by performing an ELISA assay using an antibody that specifically binds to the polypeptide.

606qb. The method of claim 605q, wherein the polypeptide is detected using an antibody array comprising an antibody that specifically binds to the polypeptide.

606qc. The method of claim 605q, wherein the detecting step comprises:

detecting modification of a substrate by the polypeptide.

609q. The method of claim 601q, wherein the sample is selected from the group consisting of:

a blood sample, a urine sample, a serum sample, an ascites sample, a saliva sample, a cell, and a portion of tissue.

610q. The method of claim 601q, wherein the sample is a tumor sample.

611q. The method of claim 610q, wherein the tumor sample is a breast tumor sample.

FIG. 1A

SEQ ID NO:1

MKTLIAAYSGVLRGERQAEADRSQRSHGGPALSREGSGRWGTGSSILSALQDLFSVTWLNRSKVEKQLQVI
SVLQWVLSFLVLGVACSAILMYIFCTDCWLIAVLYFTWLVFDWNTPKKGGRRSQWVRNWAVWRYFRDYFPI
QLVKTHNLLTTRNYIFGYHPHGIMGLGAFCNFSTEATEVSKKFPGIRPYLATLAGNFRMPVLREYLMSGGI
CPVSRDTIDYLLSKNGSGNAIIIVVGGAAESLSSMPGKNAVTLRNRKGFVKLALRHGADLVPIYSFGENEV
YKQVIFEEGSWGRWVQKKFQKYIGFAPCIFHGRGLFSSDTWGLVPYSKPITTVVGEPITIPKLEHPTQQDI
DLYHTMYMEALVKLFDKHKTKFGLPETEVLEVN

FIG. 1B

SEQ ID NO:2

ATGAAGACCCTCATAGCCGCCTACTCCGGGGTCCTGCGCGGCGAGCGTCAGGCCGAGGCTGACCGGAGCCA
GCGCTCTCACGGAGGACCTGCGCTGTCGCGCGAGGGGTCTGGGAGATGGGGCACTGGATCCAGCATCCTCT
CCGCCCTCCAGGACCTCTTCTCTGTCACCTGGCTCAATAGGTCCAAGGTGGAAAAGCAGCTACAGGTCATC
TCAGTGCTCCAGTGGGTCCTGTCCTTCCTTGTACTGGGAGTGGCCTGCAGTGCCATCCTCATGTACATATT
CTGCACTGATTGCTGGCTCATCGCTGTGCTCTACTTCACTTGGCTGGTGTTTGACTGGAACACACCCAAGA
AAGGTGGCAGGAGGTCACAGTGGGTCCGAAACTGGGCTGTGTGGCGCTACTTTCGAGACTACTTTCCCATC
CAGCTGGTGAAGACACACAACCTGCTGACCACCAGGAACTATATCTTTGGATACCACCCCCATGGTATCAT
GGGCCTGGGTGCCTTCTGCAACTTCAGCACAGAGGCCACAGAAGTGAGCAAGAAGTTCCCAGGCATACGGC
CTTACCTGGCTACACTGGCAGGCAACTTCCGAATGCCTGTGTTGAGGGAGTACCTGATGTCTGGAGGTATC
TGCCCTGTCAGCCGGGACACCATAGACTATTTGCTTTCAAAGAATGGGAGTGGCAATGCTATCATCATCGT
GGTCGGGGGTGCGGCTGAGTCTCTGAGCTCCATGCCTGGCAAGAATGCAGTCACCCTGCGGAACCGCAAGG
GCTTTGTGAAACTGGCCCTGCGTCATGGAGCTGACCTGGTTCCCATCTACTCCTTTGGAGAGAATGAAGTG
TACAAGCAGGTGATCTTCGAGGAGGGCTCCTGGGGCCGATGGGTCCAGAAGAAGTTCCAGAAATACATTGG
TTTCGCCCCATGCATCTTCCATGGTCGAGGCCTCTTCTCCTCCGACACCTGGGGGCTGGTGCCCTACTCCA
AGCCCATCACCACTGTTGTGGGAGAGCCCATCACCATCCCCAAGCTGGAGCACCCAACCCAGCAAGACATC
GACCTGTACCACACCATGTACATGGAGGCCCTGGTGAAGCTTTTCGACAAGCACAAGACCAAGTTCGGCCT
CCCGGAGACTGAGGTCCTGGAGGTGAACTGA

FIG. 1C

SEQ ID NO:3

TCCGGGACGCCAGCGCCGCGGCTGCCGCCTCTGCTGGGGTCTAGGCTGTTTCTCTCGCGCCACCACTGGCC
GCCGGCCGCAGCTCCAGGTGTCCTAGCCGCCCAGCCTCGACGCCGTCCCGGGACCCCTGTGCTCTGCGCGA
AGCCCTGGCCCCGGGGGCCGGGGCATGGGCCAGGGCGCGGGGTGAAGCGGCTTCCCGCGGGGCCGTGACTG
GGCGGGCTTCAGCCATGAAGACCCTCATAGCCGCCTACTCCGGGGTCCTGCGCGGCGAGCGTCAGGCCGAG
GCTGACCGGAGCCAGCGCTCTCACGGAGGACCTGCGCTGTCGCGCGAGGGGTCTGGGAGATGGGGCACTGG
ATCCAGCATCCTCTCCGCCCTCCAGGACCTCTTCTCTGTCACCTGGCTCAATAGGTCCAAGGTGGAAAAGC
AGCTACAGGTCATCTCAGTGCTCCAGTGGGTCCTGTCCTTCCTTGTACTGGGAGTGGCCTGCAGTGCCATC
CTCATGTACATATTCTGCACTGATTGCTGGCTCATCGCTGTGCTCTACTTCACTTGGCTGGTGTTTGACTG
GAACACACCCAAGAAAGGTGGCAGGAGGTCACAGTGGGTCCGAAACTGGGCTGTGTGGCGCTACTTTCGAG
ACTACTTTCCCATCCAGCTGGTGAAGACACACAACCTGCTGACCACCAGGAACTATATCTTTGGATACCAC
CCCCATGGTATCATGGGCCTGGGTGCCTTCTGCAACTTCAGCACAGAGGCCACAGAAGTGAGCAAGAAGTT
CCCAGGCATACGGCCTTACCTGGCTACACTGGCAGGCAACTTCCGAATGCCTGTGTTGAGGGAGTACCTGA
TGTCTGGAGGTATCTGCCCTGTCAGCCGGGACACCATAGACTATTTGCTTTCAAAGAATGGGAGTGGCAAT
GCTATCATCATCGTGGTCGGGGGTGCGGCTGAGTCTCTGAGCTCCATGCCTGGCAAGAATGCAGTCACCCT
GCGGAACCGCAAGGGCTTTGTGAAACTGGCCCTGCGTCATGGAGCTGACCTGGTTCCCATCTACTCCTTTG
GAGAGAATGAAGTGTACAAGCAGGTGATCTTCGAGGAGGGCTCCTGGGGCCGATGGGTCCAGAAGAAGTTC
CAGAAATACATTGGTTTCGCCCCATGCATCTTCCATGGTCGAGGCCTCTTCTCCTCCGACACCTGGGGGCT
GGTGCCCTACTCCAAGCCCATCACCACTGTTGTGGGAGAGCCCATCACCATCCCCAAGCTGGAGCACCCAA

TO FIG.1C-2

FIG. 1C-2

FROM FIG.1C-1

CCCAGCAAGACATCGACCTGTACCACACCATGTACATGGAGGCCCTGGTGAAGCTTTTCGACAAGCACAAG
ACCAAGTTCGGCCTCCCGGAGACTGAGGTCCTGGAGGTGAACTGAGCCAGCCTTCGGGGCCAATTCCCCTG
GAGGAACCAGCTGCAAATCACTTTTTTGCTCTGTAAATTTGGAAGTGTCATGGGTGTCTGTGGGTTATTTA
AAAGAAATTATAACAATTTTGCTAAACCATTACAATGTTAGGTCTTTTTTTAAGAAGGAAAAAGTCAGTATT
TCAAGTTCTTTCACTTCCAGCTTGCCCTGTTCTAGGTGGTGGCTAAATCTGGGCCTAATCTGGGTGGCTCA
GCTAACCTCTCTTCTTCCCTTCCTGAAGTGACAAAGGAAACTCAGTCTTCTTGGGGAAGAAGGATTGCCAT
TAGTGACTTGGACCAGTTAGATGATTCACTTTTTGCCCCTAGGGATGAGAGGCGAAAGCCACTTCTCATAC
AAGCCCCTTTATTGCCACTACCCCACGCTCGTCTAGTCCTGAAACTGCAGGACCAGTTTCTCTGCCAAGGG
GAGGAGTTGGAGAGCACAGTTGCCCCGTTGTGTGAGGGCAGTAGTAGGCATCTGGAATGCTCCAGTTTGAT
YTCCCTTCTGCCACCCCTACCTCACCCCTAGTCACTCATATCGGAGCCTGGGACTGGGCCTCCAGGATGAG
GATGGGGGTGGCAATGACACCCTGCAGGGGAAAGGACTGCCCCCCATGCACCATTGCAGGGAGGATGCCGC
CACCATGAAGCTAGGTGGAGTAACTGGTTTTTCTTGGGTGGCTGATGACATGGATGCAGCACAGACTCAGC
CTTGGCCTGGAGCACATGCTTACTGGTGGCCTCAGTTTACCTTCCCCAGATCCTAGATTCTGGATGTGAGG
AAGAGATCCCTCTTCAGAAGGGGCCTGGCCTTCTGAGCAGCAGATTAGTTCCAAAGCAGGTGGCCCCCGAA
CCCAAGCCTCACTTTTYTGTGCCTTCCTGAGGGGGTTGGGCCGGGGAGGAAACCCAACCCTCTCCTGTGTG
TTCTGTTATCTYTTGATGAGATCATTGCACCATGTCAGACTTTTGTATATGCCTTGAAAATAAATGAAAGT
GAGAATCAAAAAAAAAAAAAAAAAAAAAA

FIG. 1D-1

SEQ ID NO:4

TCCGGGACGCCAGCGCCGCGGCTGCCGCCTCTGCTGGGGTCTAGGCTGTTTCTCTCGCGCCACCACTGGCC
GCCGGCCGCAGCTCCAGGTGTCCTAGCCGCCCAGCCTCGACGCCGTCCCGGGACCCCTGTGCTCTGCGCGA
AGCCCTGGCCCCGGGGGCCGGGGCATGGGCCAGGGCGCGGGGTGAAGCGGCTTCCCGCGGGGCCGTGACTG
GGCGGGCTTCAGCCATGAAGACCCTCATAGCCGCCTACTCCGGGGTCCTGCGCGGCGAGCGTCAGGCCGAG
GCTGACCGGAGCCAGCGCTCTCACGGAGGACCTGCGCTGTCGCGCGAGGGGTCTGGGAGATGGGGCACTGG
ATCCAGCATCCTCTCCGCCCTCCAGGACCTCTTCTCTGTCACCTGGCTCAATAGGTCCAAGGTGGAAAAGC
AGCTACAGGTCATCTCAGTGCTCCAGTGGGTCCTGTCCTTCCTTGTACTGGGAGTGGCCTGCAGTGCCATC
CTCATGTACATATTCTGCACTGATTGCTGGCTCATCGCTGTGCTCTACTTCACTTGGCTGGTGTTTGACTG
GAACACACCCAAGAAAGGTGGCAGGAGGTCACAGTGGGTCCGAAACTGGGCTGTGTGGCGCTACTTTCGAG
ACTACTTTCCCATCCAGCTGGTGAAGACACACAACCTGCTGACCACCAGGAACTATATCTTTGGATACCAC
CCCCATGGTATCATGGGCCTGGGTGCCTTCTGCAACTTCAGCACAGAGGCCACAGAAGTGAGCAAGAAGTT
CCCAGGCATACGGCCTTACCTGGCTACACTGGCAGGCAACTTCCGAATGCCTGTGTTGAGGGAGTACCTGA
TGTCTGGAGGTATCTGCCCTGTCAGCCGGGACACCATAGACTATTTGCTTTCAAAGAATGGGAGTGGCAAT
GCTATCATCATCGTGGTCGGGGGTGCGGCTGAGTCTCTGAGCTCCATGCCTGGCAAGAATGCAGTCACCCT
GCGGAACCGCAAGGGCTTTGTGAAACTGGCCCTGCGTCATGGAGCTGACCTGGTTCCCATCTACTCCTTTG
GAGAGAATGAAGTGTACAAGCAGGTGATCTTCGAGGAGGGCTCCTGGGGCCGATGGGTCCAGAAGAAGTTC

TO FIG.1D-2

FIG. 1D-2

FROM FIG. 1D-1

CAGAAATACATTGGTTTCGCCCCATGCATCTTCCATGGTCGAGGCCTCTTCTCCTCCGACACCTGGGGGCT
GGTGCCCTACTCCTCCAAGCCCATCACCACTGTTGTGGGAGAGCCCCATCACCATCCCCAAGCTGGAGCACCCAA
CCCAGCAAGACATCGACCTGTACCACACCATGTACATGGAGGCCCTGGTGAAGCTTTTCGACAAGCACAAG
ACCAAGTTCGGCCTCCCCGGAGACTGAGGTCCTGGAGGTGAACTGAGCCAGCCTTCGGGGCCAATTCCCCTG
GAGGAACCAGCTGCAAATCACTTTTTTGCTCTGTAAATTTGGAAGTGTCATGGGTGTCTGTGGGTTATTTA
AAAGAAATTATATAACAATTTTGCTAAAACCATTAAAAAAAAAAAAAAAAAAAA

FIG.2-1

SEQ ID NO:5

TGGCAATTATGAAAAACTTCCCAAGGCCTACCCCTTGGACTAGCCTCTTTAAATACTTCATGTGGTCTCCA
AAATGACCCCGANTGCGATACCATTCCCATTGTGTNTCTATAAAAACCAGGGCACAGGGGANCATGCAAAC
AGCCCAAANTTACAAANCCATGAGGTGGAGGGCTAGGATCTGAACCCAGNTCTGTCTGATTCTATANCTGA
TGCTCTTCTCATATCTAAAAGGGTACCTGTGGGAGGTGAGGTTTGTACTGGGGACCCCATGACTGGAAAAA
AGGGTGACAGTGGACTGACATCTTCCCTCTGCTGTAGGCACTGGATCCANCATCCTCTCCGCCCTCCANAA
CCTCTTCTCTNTCACCTGGCTCAATAGGTCCAAGGTGGAAAAGCANCTACAGGTCATCTCAGTGCTCCAGT
GGGTCCTGTCCTTCCTTGTACTGGGANTGGCCTGCAGTGCCATCCTCATGTACATATTCTGCACTGATTGC
TGGCTCATCGCTGTGCTCTACTTCACTTGGCTGGTGTTTGACTGGAACACACCCAAGAAAGGTGGCAGGAG
GTCACAGTGGGTCCGAAACTGGGCTGTGTGGCGCTACTTTCGAGACTACTTTCCCATCCAGCTGGTGAAGA
CACACAACCTGCTGACCACCAGGAACTATATCTTTGGATACCACCCCCATGGTATCATGGGCCTGGGTGCC
TTCTGCAACTTCAGCACAGAGGCCACAGAAGTGAGCAAGAAGTTCCCAGGCATACGGCCTTACCTGGCTAC
ACTGGCAGGCAACTTCCGAATGCCTGTGTTGAGGGAGTACCTGATGTCTGGAGGTATCTGCCCTGTCAGCC
GGGACACCATAGACTATTTGCTTTCAAAGAATGGGAGTGGCAATGCTATCATCATCGTGGTCGGGGGTGCG
GCTGAGTCTCTGAGCTCCATGCCTGGCAAGAATGCAGTCACCCTGCGGAACCGCAAGGGCTTTGTGAAACT
GGCCCTGCGTCATGGAGCTGACCTGGTTCCCATCTACTCCTTTGGAGAGAATGAAGTGTACAAGCAGGTGA
TCTTCGAGGAGGGCTCCTGGGGCCGATGGGTCCAGAAGAAGTTCCAGAAATACATTGGTTTCGCCCCATGC
ATCTTCCATGGTCGAGGCCTCTTCTCCTCCGACACCTGGGGGCTGGTGCCCTACTCCAAGCCCATCACCAC
TGTTGTGGGAGAGCCCATCACCATCCCCAAGCTGGAGCACCCAACCCAGCAAGACATCGACCTGTACCACA

TO FIG.2-2

FIG.2-2

FROM FIG.2-1

CCATGTACATGGAGGCCCTGGTGAAGCTTTTCGACAAGCACAAGACCAAGTTCGGCCTCCCGGAGACTGAG
GTCCTGGAGGTGAACTGAGCCAGCCTTCGGGGCCAATTCCCCTGGAGGAACCAGCTGCAAATCACTTTTTT
GCTCTGTAAATTTGGAAGTGTCATGGGTGTCTGTGGGTTATTTAAAAGAAATTATAACAATTTTGCTAAAC
CATTAAAAAAAAAAAAAAAAAAAAAARAARRAAAAAGTCAGTATTTCAAGTTCTTTCACTTCCAGCTTGCCC
TGTTCTAGGTGGTGGCTAAATCTGGGCCTAATCTGGGTGGCTCAGCTAACCTCTCTTCTTCCCTTCCTGAA
GTGACAAAGGAAACTCAGTCTTCTTGGGGAAGAAGGATTGCCATTAGTGACTTGGACCAGTTAGATGATTC
ACTTTTTGCCCCTAGGGATGAGAGGCGAAAGCCACTTCTCATACAAGCCCCTTTATTGCCACTACCCCACG
CTCGTCTAGTCCTGAAACTGCAGGACCAGTTTCTCTGCCAAGGGGAGGAGTTGGAGAGCACAGTTGCCCCG
TTGTGTGAGGGCAGTAGTAGGCATCTGGAATGCTCCAGTTTGATYTCCCTTCTGCCACCCCTACCTCACCC
CTAGTCACTCATATCGGAGCCTGGGACTGGGCCTCCAGGATGAGGATGGGGGTGGCAATGACACCCTGCAG
GGGAAAGGACTGCCCCCCATGCACCATTGCAGGGAGGATGCCGCCACCATGAAGCTAGGTGGAGTAACTGG
TTTTTCTTGGGTGGCTGATGACATGGATGCAGCACAGACTCAGCCTTGGCCTGGAGCACATGCTTACTGGT
GGCCTCAGTTTACCTTCCCCAGATCCTAGATTCTGGATGTGAGGAAGAGATCCCTCTTCAGAAGGGGCCTG
GCCTTCTGAGCAGCAGATTAGTTCCAAAGCAGGTGGCCCCCGAACCCAAGCCTCACTTTTYTGTGCCTTCC
TGAGGGGGTTGGGCCGGGGAGGAAACCCAACCCTCTCCTGTGTGTTCTGTTATCTYTTGATGAGATCATTG
CACCATGTCAGACTTTTGTATATGCCTTGAAAATAAATGAAAGTGAGAATCAAAAAAAAAAAAAAAAAAAA
AAAA

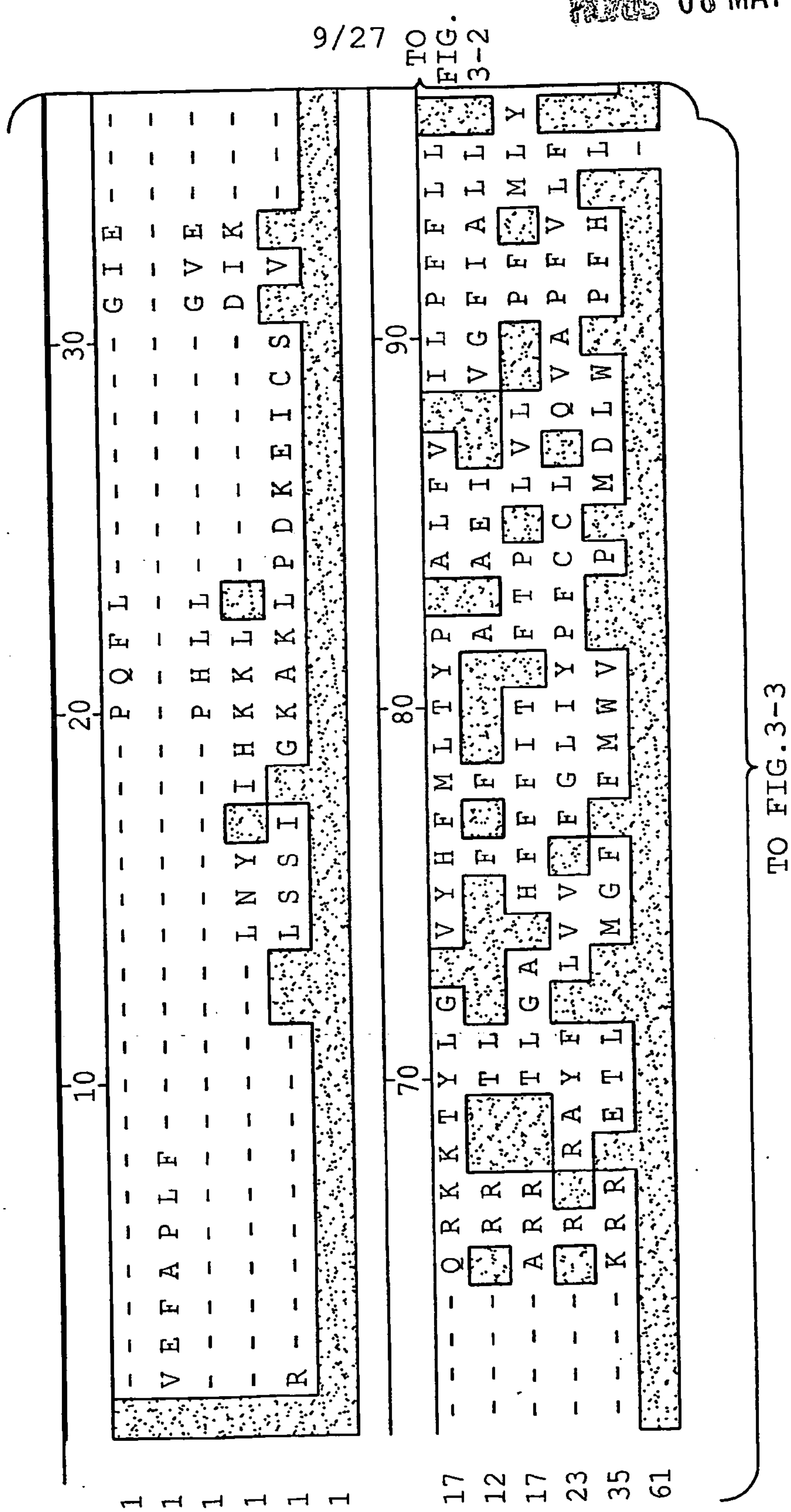
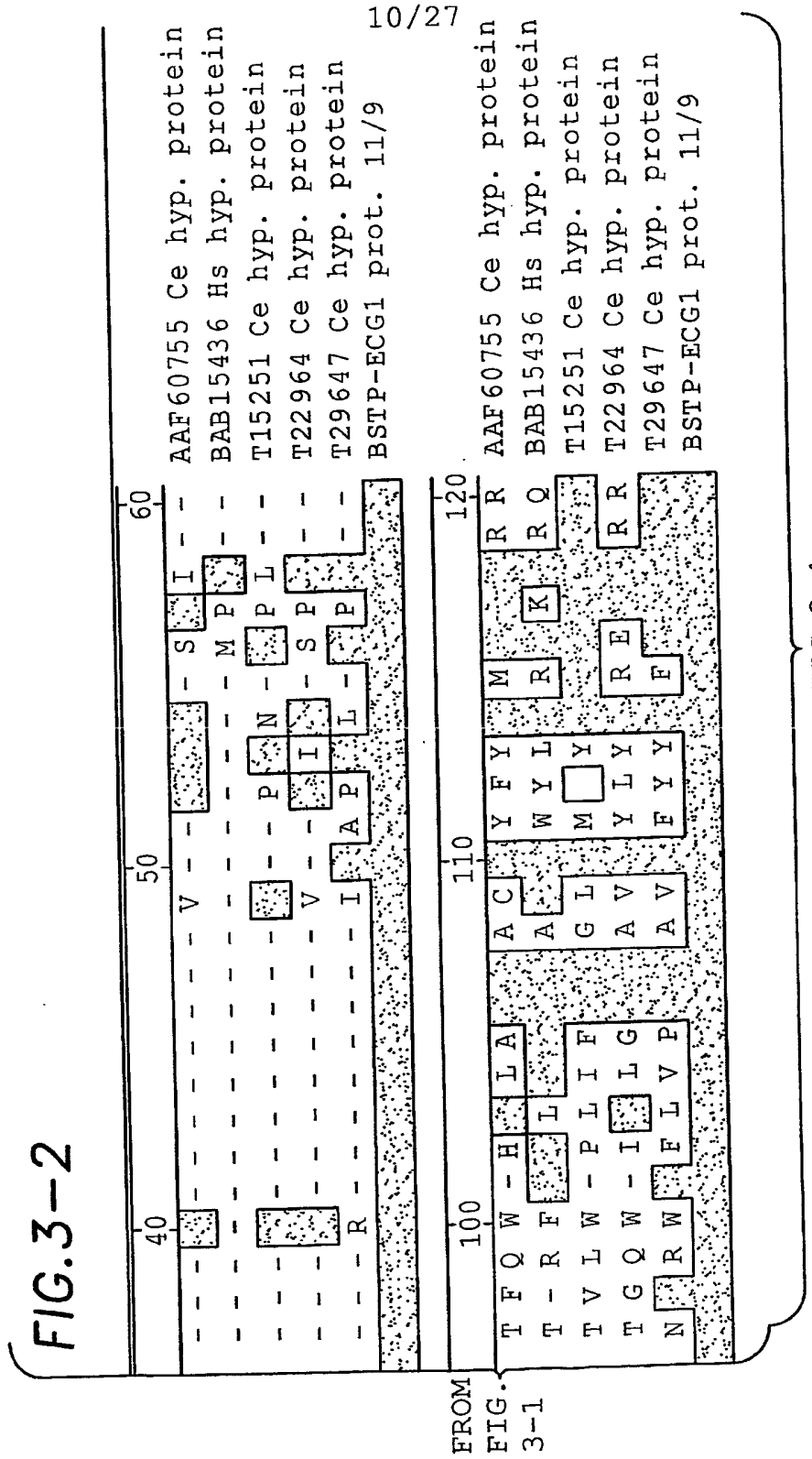


FIG.3-1

FIG.3-2


40
50
60
100
110
120
AAF60755 Ce hyp. protein
BAB15436 Hs hyp. protein
T15251 Ce hyp. protein
T22964 Ce hyp. protein
T29647 Ce hyp. protein
BSTP-ECG1 prot. 11/9
FROM FIG. 3-1
TO FIG. 3-4

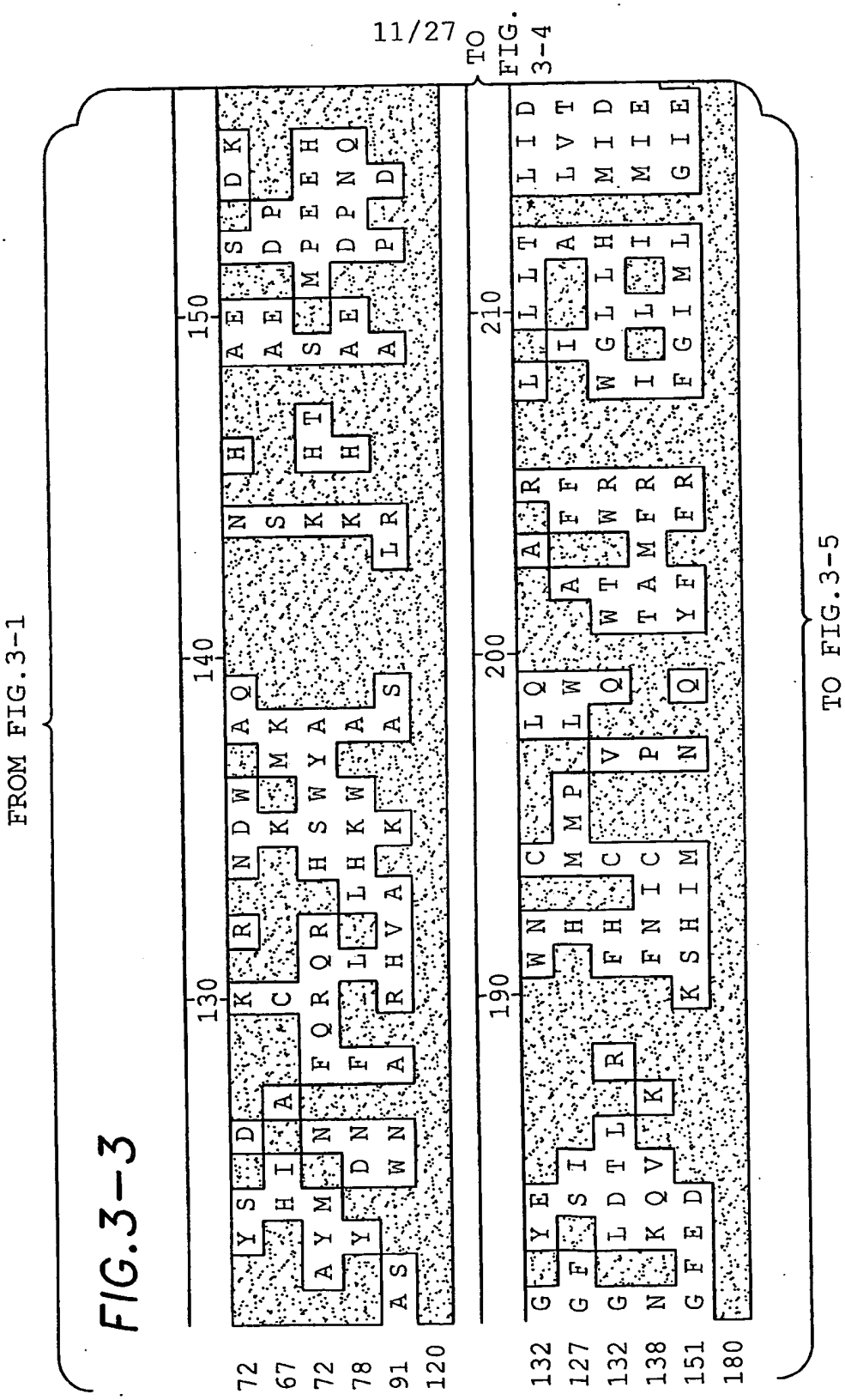

TO FIG. 3-4
FROM FIG. 3-1
TO FIG. 3-5
130
140
150
190
200
210
72
67
72
78
91
120
132
127
132
138
151
180

FIG. 3-3

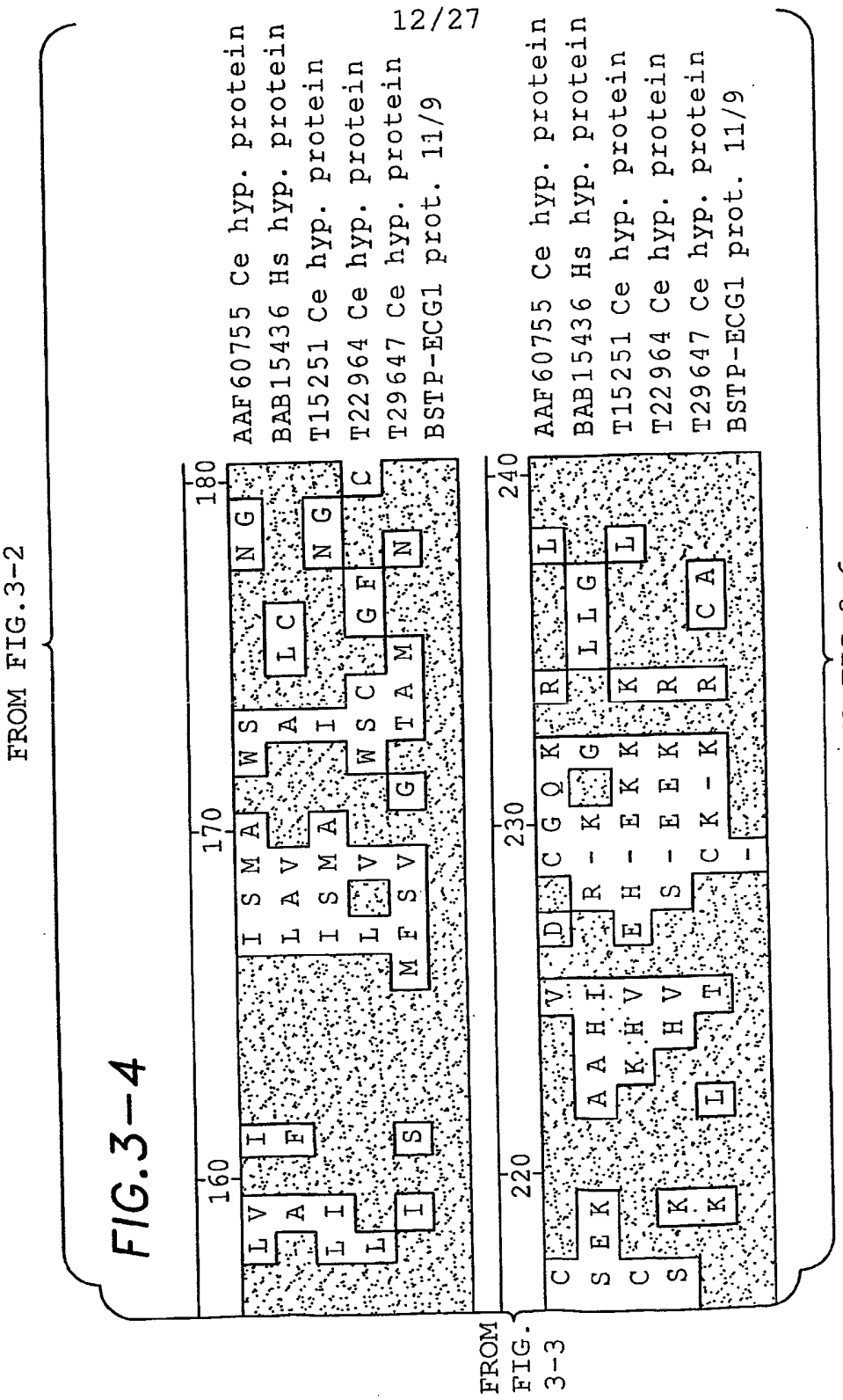

FROM FIG.3-2
TO FIG.3-6
FROM FIG. 3-3
160
170
180
220
230
240
AAF60755 Ce hyp. protein
BAB15436 Hs hyp. protein
T15251 Ce hyp. protein
T22964 Ce hyp. protein
T29647 Ce hyp. protein
BSTP-ECG1 prot. 11/9

FIG.3-4

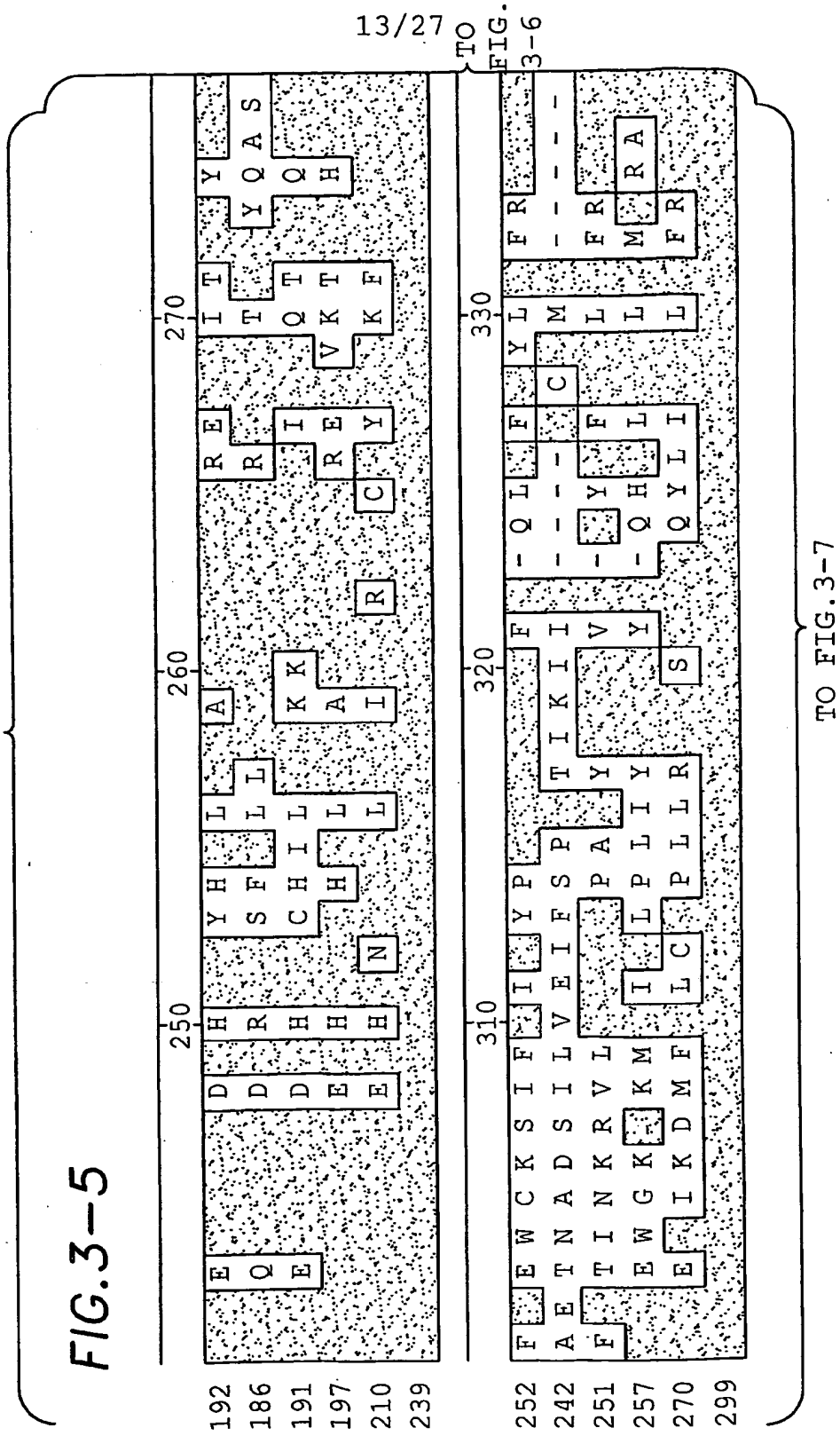

TO FIG. 3-6
FROM FIG. 3-3
TO FIG. 3-7
250
260
270
310
320
330
192
186
191
197
210
239
252
242
251
257
270
299

FIG. 3-5

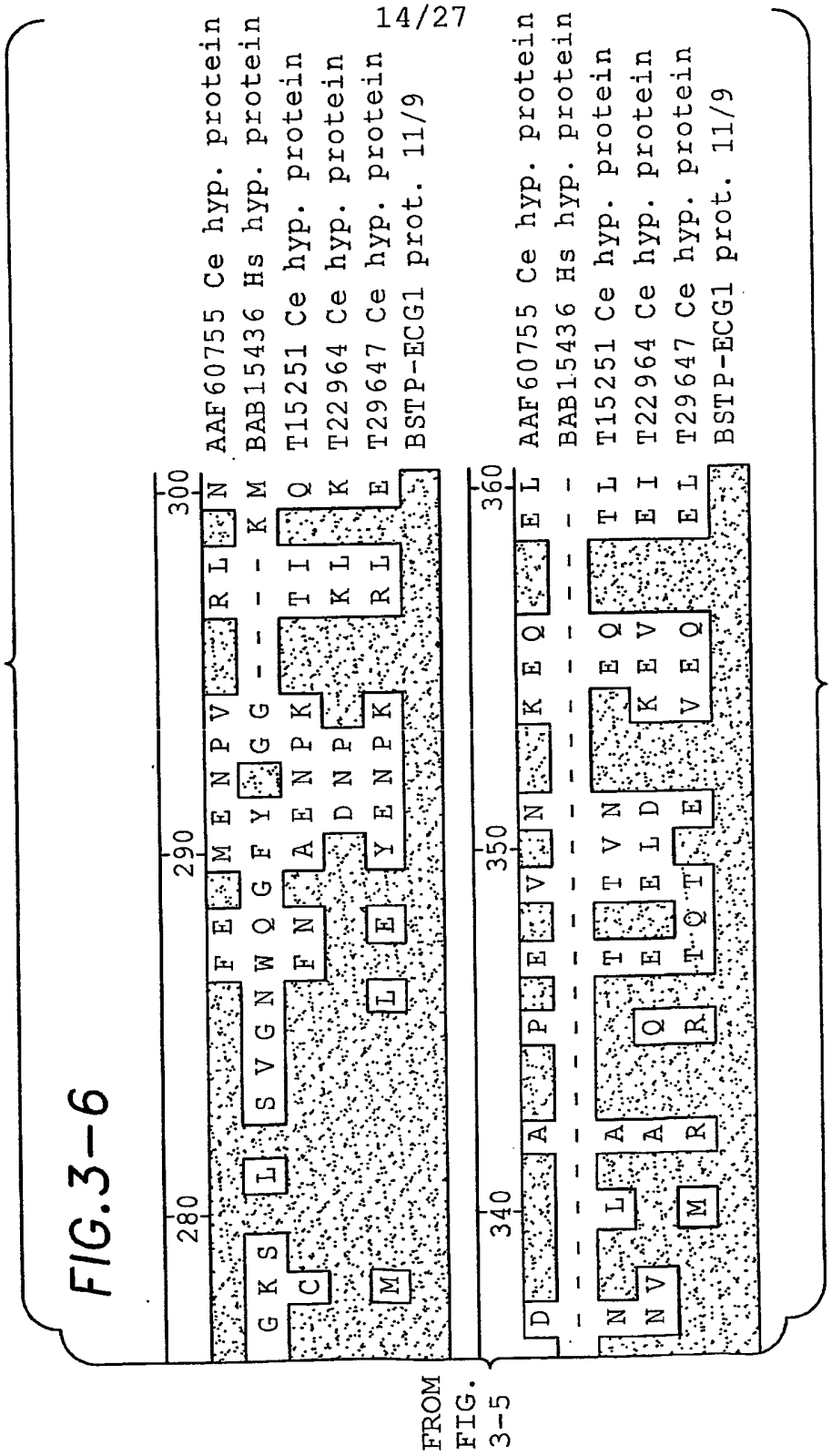


FIG.3-6
FROM FIG.3-4
TO FIG.3-8
FROM FIG. 3-5
280
290
300
340
350
360
AAF60755 Ce hyp. protein
BAB15436 Hs hyp. protein
T15251 Ce hyp. protein
T22964 Ce hyp. protein
T29647 Ce hyp. protein
BSTP-ECG1 prot. 11/9

TO FIG. 3-8

FROM FIG. 3-5

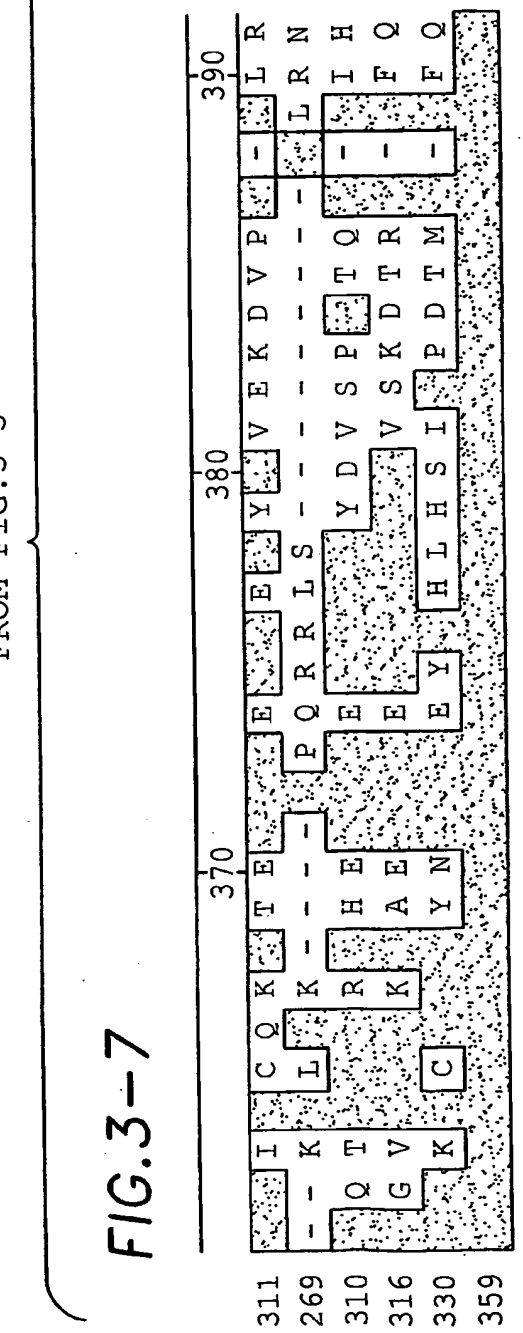


FIG. 3-7

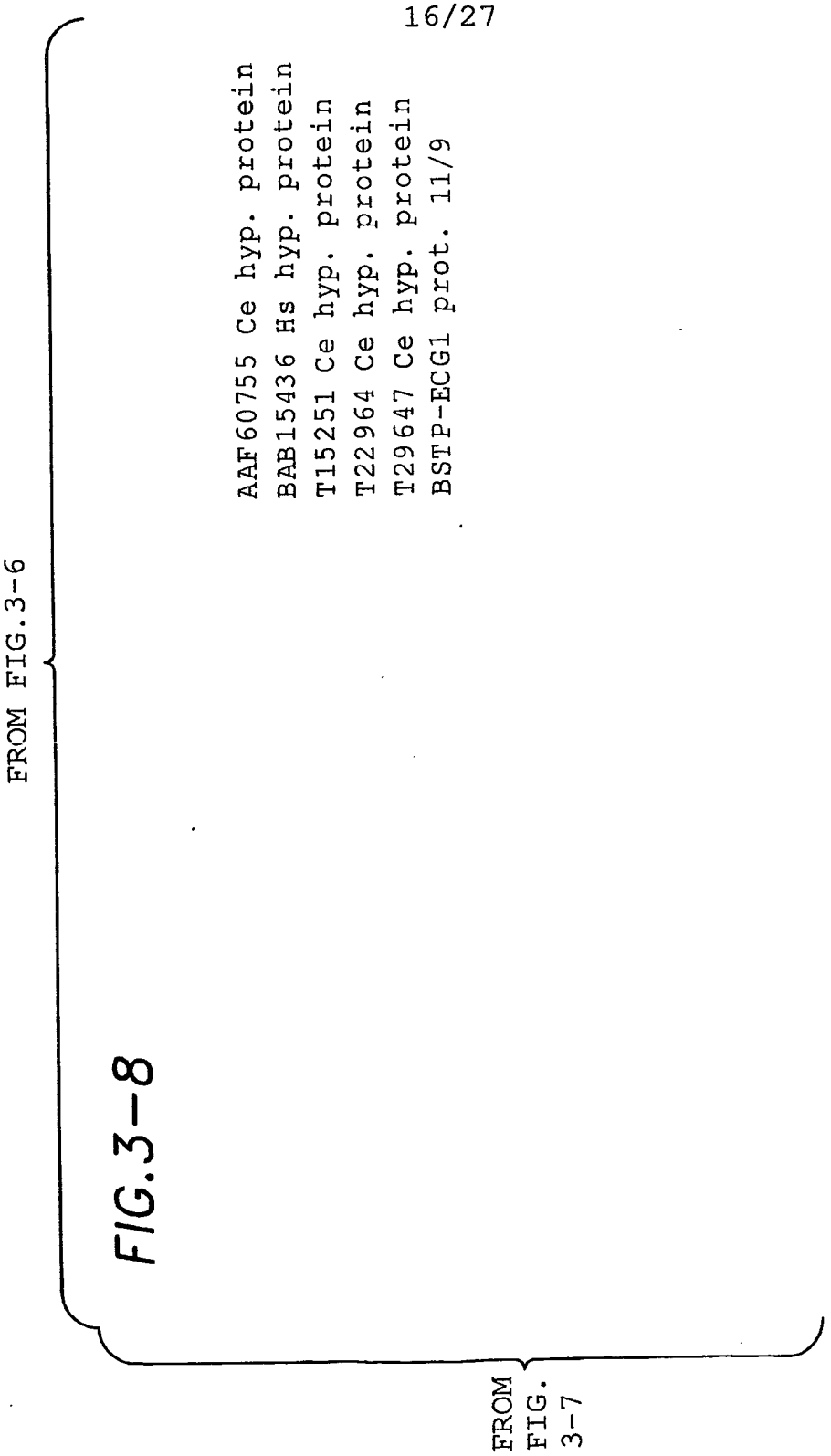

FROM FIG.3-6
AAF60755 Ce hyp. protein
BAB15436 Hs hyp. protein
T15251 Ce hyp. protein
T22964 Ce hyp. protein
T29647 Ce hyp. protein
BSTP-ECG1 prot. 11/9
FROM FIG. 3-7

FIG.3-8

FIG. 4-1

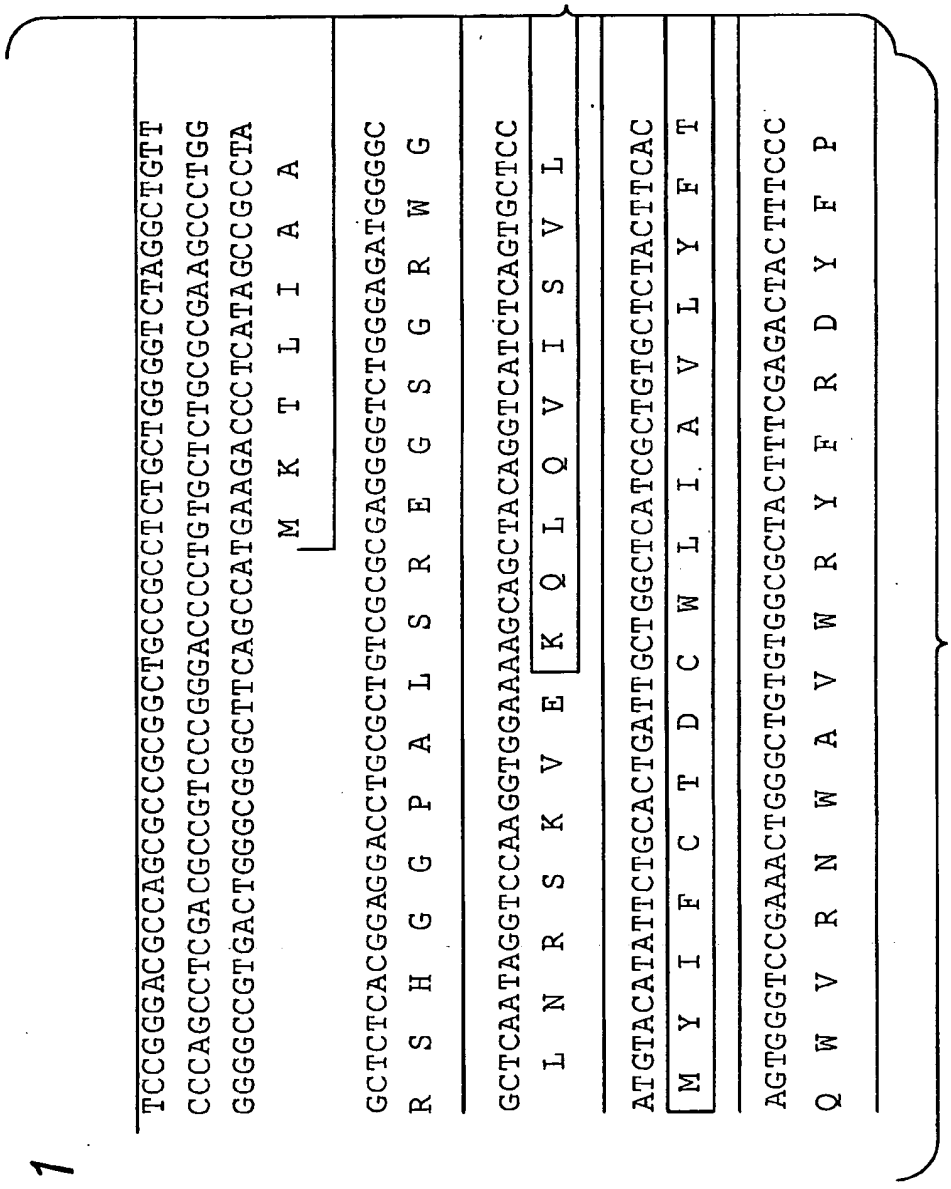

TCCGGGACGCCAGCGCCGCGGCTGCCGCCTCTGCTGGGGTCTAGGCTGTT
CCCAGCCCTCGACGCCGTCCCGGGACCCCCTGTGCTCTGCGCGAAGCCCTGG
GGGGCCGTGACTGGGCGGGCTTCAGCCATGAAGACCCTCATAGCCGCCTA
M K T L I A A
GCTCTCACGGAGGACCTGCGCTGTCGCGCGAGGGGTCTGGGAGATGGGGC
R S H G G G P A L S R R E G G S G G R W G
TO FIG. 4-2
GCTCAATAGGTCCAAGGTGGAAAAGCAGCTACAGGTCATCTCAGTGCTCC
L N R S K V E K Q L Q V I S V L
ATGTACATATTCTGCACTGATTGCTGGCTCATCGCTGTGCTCTACTTCAC
M Y I F C T D C W L I A V L Y F T
AGTGGGTCCGAAACTGGGCTGTGTGGCGCTACTTTCGAGACTACTTTCCC
Q W V R N W A V W R Y F R D Y F P
TO FIG. 4-3

FIG. 4-2

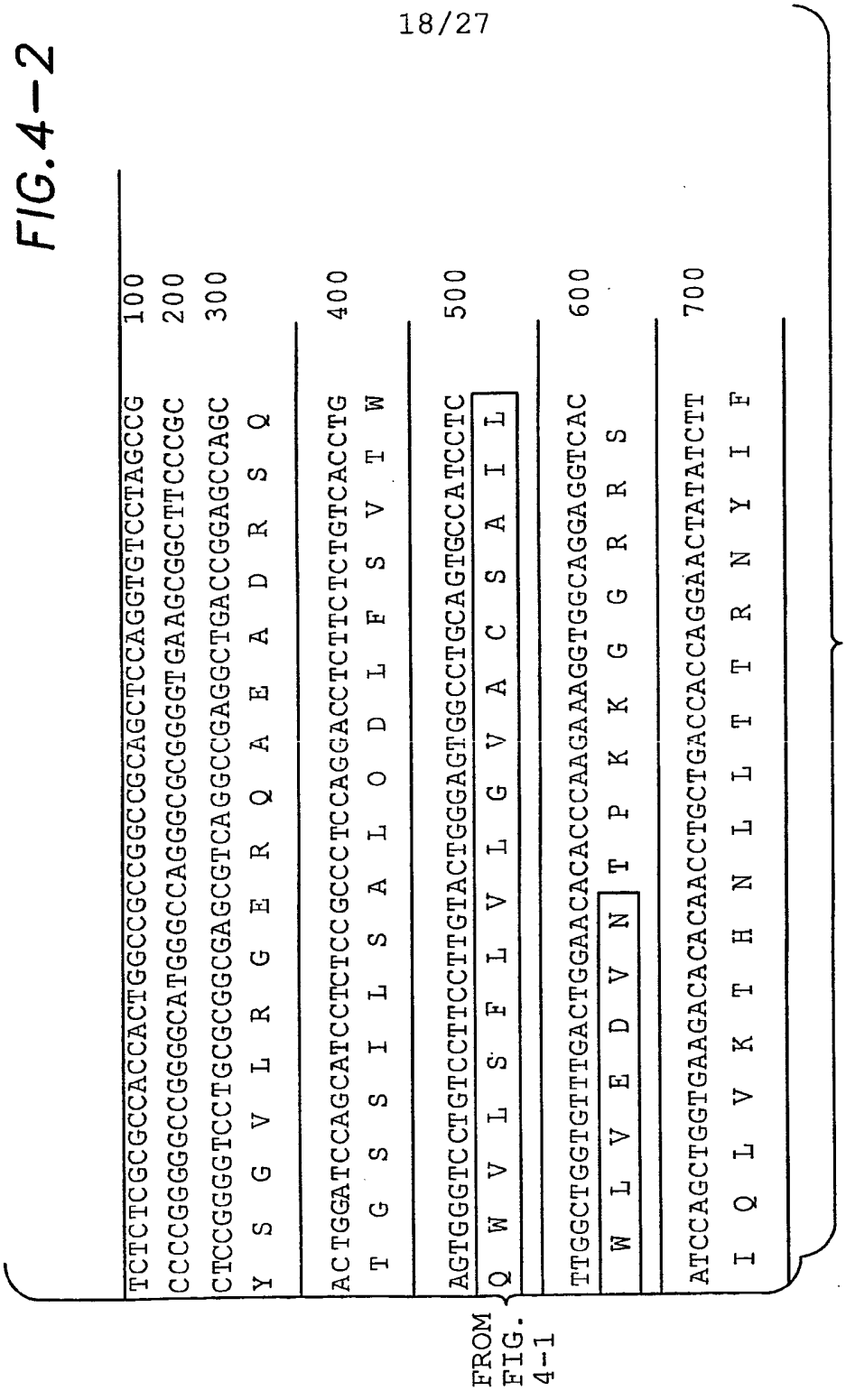
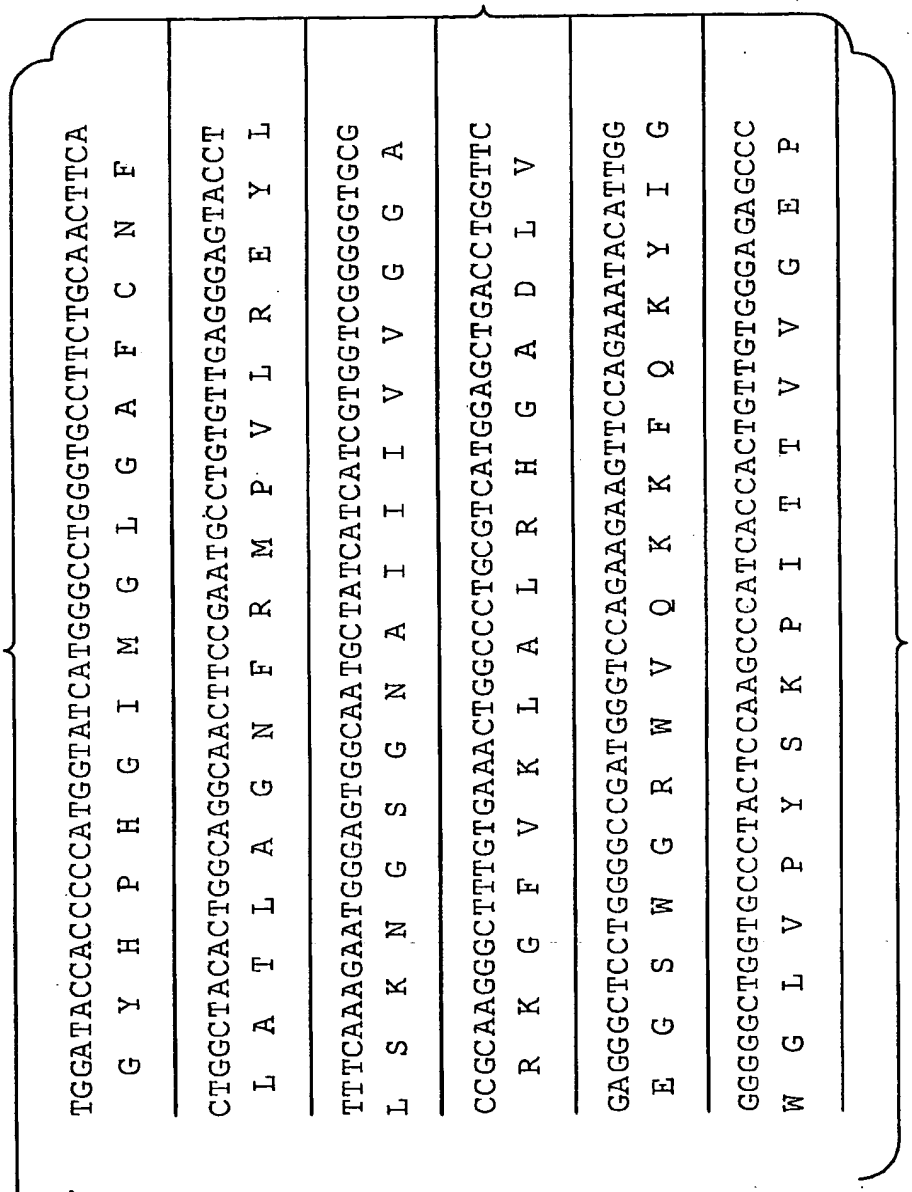

TCTCTCGCGCCACCACTGGCCGCCGGCCGCAGCTCCAGGTGTCCTAGCCG 100
CCCCGGGGGCCGGGGCATGGGCCAGGGCGCGGGGGTGAAGCGGCTTCCCGC 200
CTCCGGGGTCCTGCGCGGCGAGCGTCAGGCCGAGGCTGACCGGAGCCAGC 300
Y S G V L R G E R Q A E A D R S Q
ACTGGATCCAGCATCCTCTCCGCCCTCCAGGACCTCTTCTCTGTCACCTG 400
T G S S I L S A L O D D L F S V T W
AGTGGGTCCTGTCCTTCCTTGTACTGGGAGTGGCCTGCAGTGCCATCCTC 500
Q W V L S F L V L G V A C S A I L
TTGGCTGGTGTTTGACTGGAACACACCCAAGAAAGGTGGCAGGAGGTCAC 600
W L V E D V N T P K K G G R R S
ATCCAGCTGGTGAAGACACACAACCTGCTGACCACCAGGAACTATATCTT 700
I Q L V K T H N L L T T R N Y I F
FROM FIG. 4-1
TO FIG. 4-4

FIG. 4-3


FROM FIG. 4-1
TO FIG. 4-4
TO FIG. 4-5
TGGATACCACCCCCATGGTATCATGGGCCTGGGTGCCTTCTGCAACTTCA
G Y H P H G I M G L G A F C N F
CTGGCTACACTGGCAGGCAACTTCCGAATGCCTGTGTTGAGGGAGTACCT
L A T L A G N F R M P V L R E Y L
TTTCAAAGAATGGGAGTGGCAATGCTATCATCATCGTGGTCGGGGGTGCG
L S K N G S G N A I I I V V G G A
CCGCAAGGGCTTTGTGAAACTGGCCCTGCGTCATGGAGCTGACCTGGTTC
R K G F V K L A L R H G A D L V
GAGGGCTCCTGGGGCCGATGGGTCCAGAAGAAGTTCCAGAAATACATTGG
E G S W G R W V Q K K F Q K Y I G
GGGGGCTGGTGCCCTACTCCAAGCCCATCACCACTGTTGTGGGAGAGCCC
W G L V P Y S K P I T T V V G E P

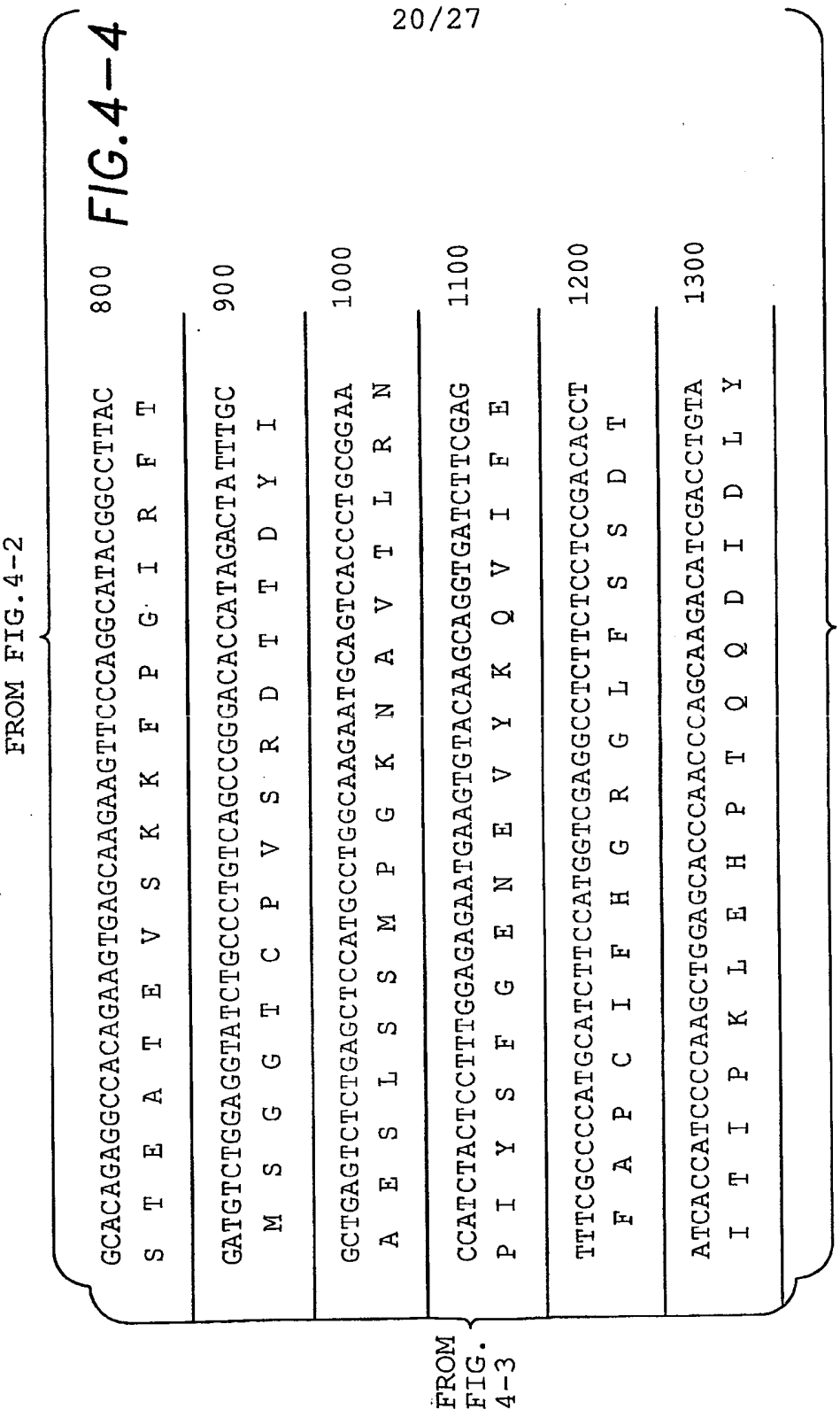

FROM FIG. 4-2
GCACAGAGGCCACAGAAGTGAGCAAGAAGTTCCCAGGCATACGGCCTTAC
S T E A T E V S K K F P G I R F T
800
GATGTCTGGAGGTATCTGCCCTGTCAGCCGGGACACCATAGACTATTTGC
M S G G T C P V S R D T T D Y I
900
GCTGAGTCTCTGAGCTCCATGCCTGGCAAGAATGCAGTCACCCTGCGGAA
A E S L S S M P G K N A V T L R N
1000
CCATCTACTCCTTTGGAGAGAATGAAGTGTACAAGCAGGTGATCTTCGAG
P I Y S F G E N E V Y K Q V I F E
1100
TTTCGCCCCATGCATCTTCCATGGTCGAGGCCTCTTCTCCTCCGACACCT
F A P C I F H G R G L F S S D T
1200
ATCACCATCCCCAAGCTGGAGCACCCAACCCAGCAAGACATCGACCTGTA
I T I P K L E H P T Q Q D I D L Y
1300
FROM FIG. 4-3
TO FIG. 4-6

FIG. 4-4

FIG. 4-5

FROM FIG. 4-3

CCACACCATGTACATGGAGGCCCTGGTGAAGCTTTTCGACAAGCACAAGA

H T M Y M E A L V K L F D K H H K

CTTCGGGGCCAATTCCCCTGGAGGAACCAGCTGCAAATCACTTTTTTGCT

ATAACAATTTTGCTAAACCATTACAATGTTAGGTCTTTTTTAAGAAGGAA

GTGGCTAAATCTGGGCCTAATCTGGGTGGCTCAGCTAACCTCTCTTCTTC

CCATTAGTGACTTGGACCAGTTAGATGATTCACTTTTTGCCCCTAGGGAT

CGCTCGTCTAGTCCTGAAACTGCAGGACCAGTTTCTCTGCCAAGGGGAGG

GAATGCTCCAGTTTGATYTCCCTTCTGCCACCCCCTACCTCACCCCTAGTC

AATGACACCCTGCAGGGGAAAGGACTGCCCCCCATGCACCATTGCAGGGA

CTGATGACATGGATGCAGCACACAGACTCAGCCTTGGCCTGGAGCACATGCT

GAAGAGATCCCCTCTTCAGAAGGGGCCTGGCCTTCTGAGCAGCAGATTAGT

GAGGGGGTTGGGCCGGGGAGGAAACCCAACCCTCTCCTGTGTGTTCTGTT

AAATAAATGAAAGTGAGAATCAAAAAAAAAAAAAAAAAAAAAAAAAA 2445

TO FIG. 4-6

FROM FIG.4-4

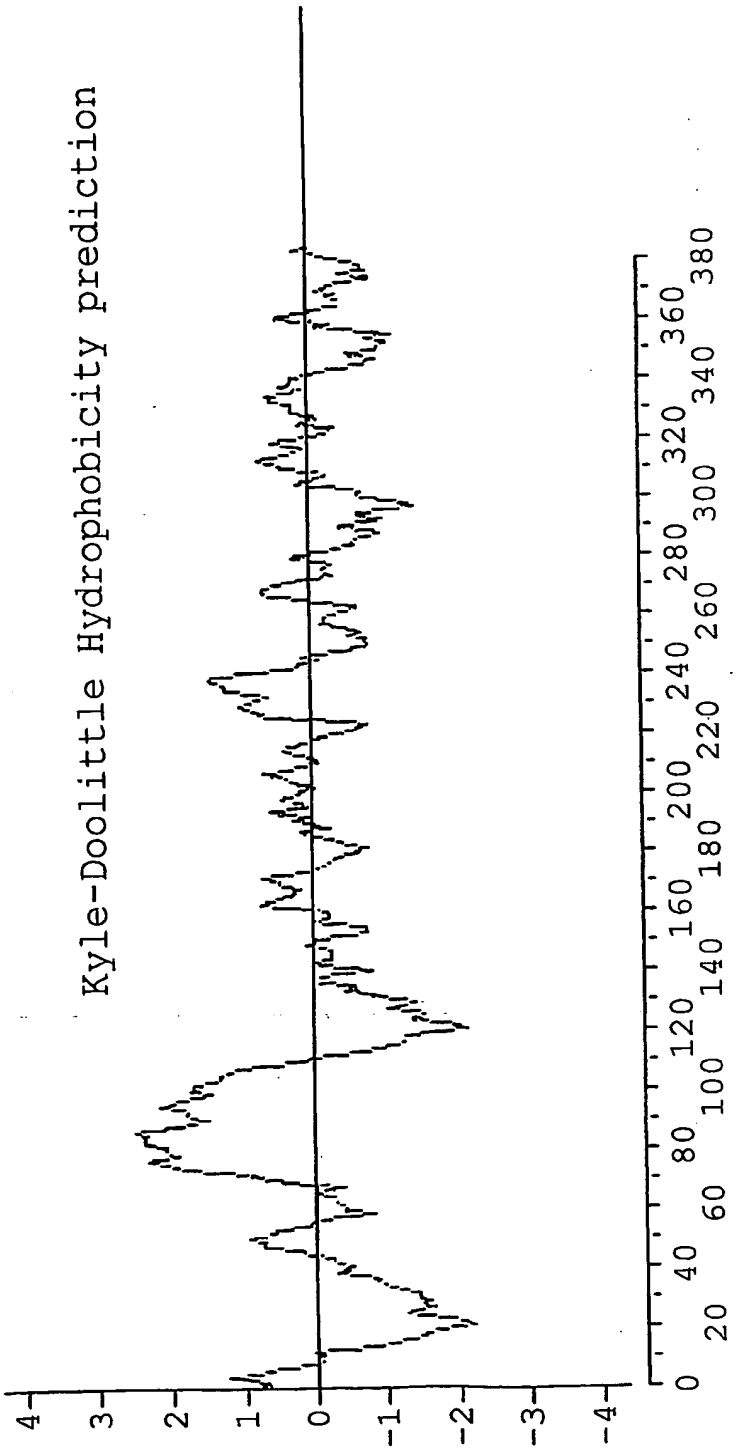
FIG.4-6

```
CCAAGTTCGGCCTCCCGGAGACTGAGGTCCTGGAGGTGAACTGAGCCAGC     1400
T  K  F  G  L  P  E  T  E  V  L  E  V  N
```

FROM FIG. 4-5

```
CTGTAAATTTGGAAGTGTCATGGGTGTCTGTGGGTTATTTAAAAGAAATT     1500
AAAGTCAGTATTTCAAGTTCTTTCACTTCCAGCTTGCCCTGTTCTAGGTG     1600
CCTTCCTGAAGTGACAAAGGAAACTCAGTCTTCTTGGGGAAGAAGGATTG     1700
GAGAGGCGAAAGCCACTTCTCATACAAGCCCCTTTATTGCCACTACCCCA     1800
AGTTGGAGAGCACAGTTGCCCCGTTGTGTGAGGGCAGTAGTAGGCATCTG     1900
ACTCATATCGGAGCCTGGGACTGGGCCTCCAGGATGAGGATGGGGGTGGC     2000
GGATGCCGCCACCATGAAGCTAGGTGGAGTAACTGGTTTTTCTTGGGTGG     2100
TACTGGTGGCCTCAGTTTACCTTCCCCAGATCCTAGATTCTGGATGTGAG     2200
TCCAAAGCAGGTGGCCCCCGAACCCAAGCCTCACTTTTYTGTGCCTTCCT     2300
ATCTYTTGATGAGATCATTGCACCATGTCAGACTTTTGTATATGCCTTGA     2400
```

FIG.5A


Kyle-Doolittle Hydrophobicity prediction
4
3
2
1
0
-1
-2
-3
-4
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
380

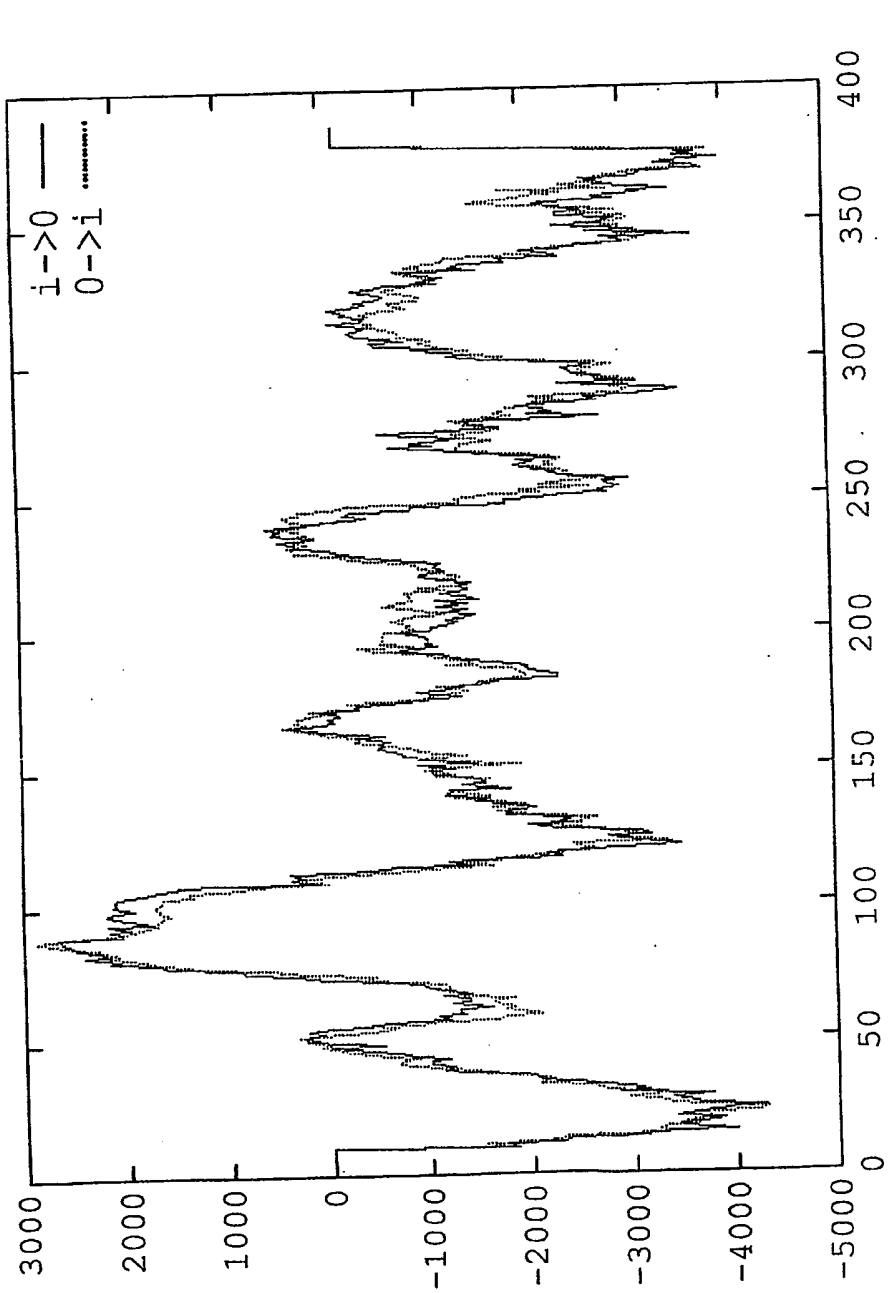

TMpred output for Transmembrane prediction for BSTP-ECG1
i->o
o->i
3000
2000
1000
0
-1000
-2000
-3000
-4000
-5000
0
50
100
150
200
250
300
350
400

FIG.5B

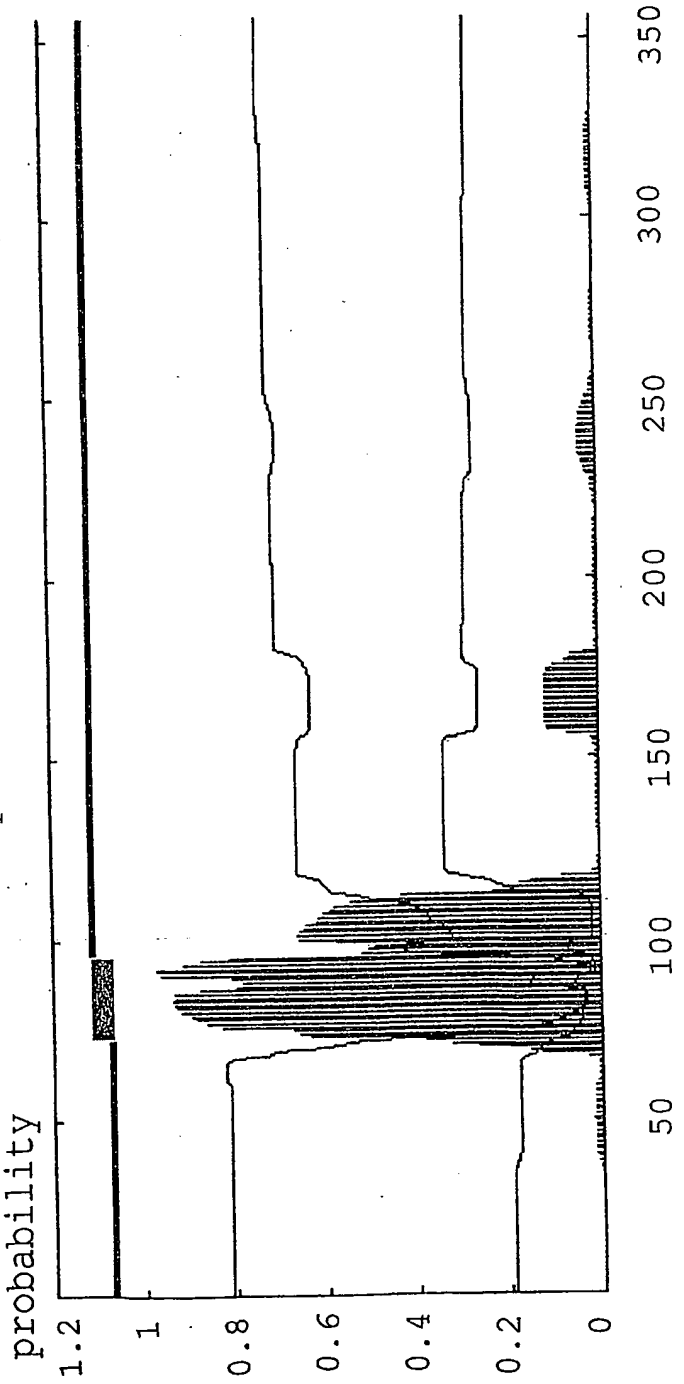

Prediction of Transmembrane helices
in proteins TMHMM (v.2.0)
probability
1.2
1
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300
350

FIG.5C

FIG. 6A

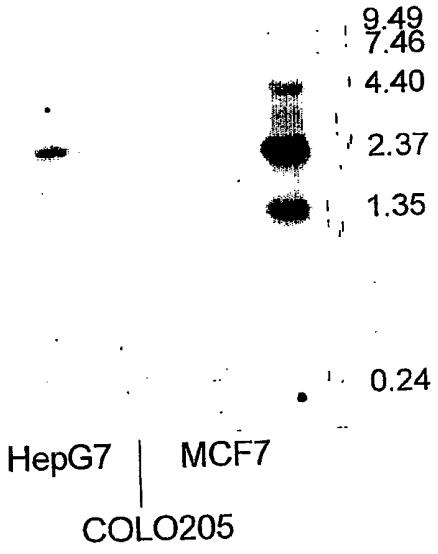
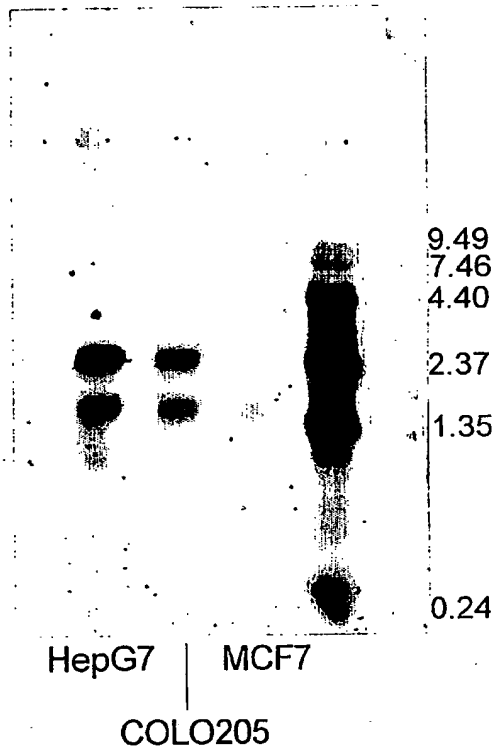


FIG. 6B



INTERNATIONAL SEARCH REPORT

Int al Application No
PCT/US 01/23439

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/52 C12N9/10 C12Q1/68 G01N33/574 C07K16/40
A61K38/43 A61K39/395 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q G01N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, EMBL, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 12708 A (BAKER KEVIN ;GENENTECH INC (US); GODDARD AUDREY (US); GURNEY AUSTI) 9 March 2000 (2000-03-09) SEQ ID NOs: 291 and 292	1-45,47, 50,53,59
X	WO 99 47655 A (SCHMITT ARMIN ;SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BERND) 23 September 1999 (1999-09-23)	2-4,21, 24-31, 34, 41-45, 47,65, 98-103
Y	SEQ ID NOs: 21 and 95 page 4, line 25 - line 42	47-64, 66-96, 104-153

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 May 2002	07/06/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Mandl, B

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int[illegible]ial Application No
PCT/US 01/23439

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MARTIN K. J. ET AL.: "LINKING GENE EXPRESSION PATTERNS TO THERAPEUTIC GROUPS IN BREAST CANCER" CANCER RESEARCH, vol. 60, no. 8, 15 April 2000 (2000-04-15), pages 2232-2238, XP001026395 ISSN: 0008-5472 abstract	47-64, 66-96, 104-153
Y	MENDOZA L. G. ET AL.: "HIGH-THROUGHPUT MICROARRAY-BASED ENZYME-LINKED IMMUNOSORBENT ASSAY (ELISA)" BIOTECHNIQUES, vol. 27, no. 4, October 1999 (1999-10), pages 778,780,782-786,788, XP000992893 ISSN: 0736-6205 page 778, middle column	53-64, 79-96, 104-153
A	WO 99 44063 A (US HEALTH ;KONONEN JUHA (US); KALLIONIEMI OLLI (US); LEIGHTON STEP) 2 September 1999 (1999-09-02) abstract; examples 1-4	
A	GOLUB T. R. ET AL.: "MOLECULAR CLASSIFICATION OF CANCER: CLASS DISCOVERY AND CLASS PREDICTION BY GENE EXPRESSION MONITORING" SCIENCE, vol. 286, 15 October 1999 (1999-10-15), pages 531-537, XP002905479 ISSN: 0036-8075 the whole document	
A	NACHT M. ET AL:: "Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer" CANCER RESEARCH, vol. 59, no. 21, 1 November 1999 (1999-11-01), pages 5464-5470, XP002166291 ISSN: 0008-5472 abstract	

-/--

INTERNATIONAL SEARCH REPORT

Int al Application No
PCT/US 01/23439

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T	CASES S. ET AL.: "Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 276, no. 42, 19 October 2001 (2001-10-19), pages 38870-38876, XP002199715 October 19, 2001 ISSN: 0021-9258 figures 2,3 -----	1-38

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 65,100,101 and claims 98,99,102 as far as an in vivo application and a method of treatment are concerned, are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition. And although claims 110,111,120-126,136-153 and claims 104-119,127-135 as far as the method enables a decision on the treatment necessary, are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Due to the lack of complience with Rule 6(1)(b) PCT, the claims have been renumbered consecutively with Arabic numerals and the documents found in the search report have been cited against the claims so renumbered.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int al Application No
PCT/US 01/23439

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 0012708 A	09-03-2000	AU	5590899 A	21-03-2000
		EP	1144629 A2	17-10-2001
		WO	0012708 A2	09-03-2000
		AU	6041399 A	10-04-2000
		EP	1115863 A1	18-07-2001
		WO	0017353 A1	30-03-2000
WO 9947655 A	23-09-1999	DE	19813835 A1	23-09-1999
		WO	9947655 A2	23-09-1999
		EP	1070125 A2	24-01-2001
		JP	2002506639 T	05-03-2002
WO 9944063 A	02-09-1999	AU	2875799 A	15-09-1999
		AU	2973599 A	15-09-1999
		CA	2318789 A1	02-09-1999
		CA	2318984 A1	02-09-1999
		EP	1068528 A2	17-01-2001
		EP	1066517 A1	10-01-2001
		JP	2002505431 T	19-02-2002
		JP	2002505432 T	19-02-2002
		WO	9944062 A1	02-09-1999
		WO	9944063 A2	02-09-1999